

Weighted Adaptive Multiple Decision Functions for False Discovery Rate Control

Joshua D. Habiger*

301G MSCS building, Stillwater, OK 74078

Abstract: Efforts to develop more efficient multiple hypothesis testing procedures for false discovery rate (FDR) control have focused on incorporating an estimate of the proportion of true null hypotheses (such procedures are called adaptive) or exploiting heterogeneity across tests via some optimal weighting scheme. This paper combines these approaches and provides an assessment of the resulting methodology. Specifically, weighted adaptive multiple decision functions (WAMDFs) that control the FDR when test statistics from true null hypotheses are independent and independent of test statistics from false null hypotheses are provided, and optimal weights for a random effects model are derived. Assessment reveals that, under a weak dependence structure, WAMDFs dominate procedures which are either adaptive or weighted but not both. In particular, they allow for more rejected null hypotheses and have false discovery proportion that is asymptotically less than or equal to the nominal level, even when the employed weights are only positively correlated with optimal weights. The method is demonstrated on a real data set.

MSC 2010 subject classifications: Primary62F03; secondary62J15.

Keywords and phrases: Adaptive, False Discovery Proportion, False Discovery Rate, Heterogeneity, Multiple Decision Function, Weighted, Weighted p -Value.

1. Introduction

High throughput technology routinely generates data sets that allow for hundreds or thousands of null hypotheses to be tested simultaneously. For example, in [Anderson and Habiger \(2012\)](#), RNA sequencing technology was used to measure the prevalence of bacteria living near the roots of wheat plants across $i = 1, 2, \dots, 5$ treatment groups for each of $m = 1, 2, \dots, M = 778$ bacteria, thereby facilitating the simultaneous testing of 778 null hypotheses. More specifically, denote the prevalence of bacteria m in the i th treatment group by Y_{im} and denote the mean of Y_{im} by μ_{im} . Assume $\log(\mu_{im}) = \beta_{0m} + \beta_{1m}x_i$, where β_{0m} and β_{1m} are regression parameters and x_i is the total shoot biomass of wheat plants in the i th group. One objective was to determine which bacteria are positively associated with shoot biomass via the testing of the null hypothesis $H_m : \beta_{1m} = 0$ against the alternative hypothesis $K_m : \beta_{1m} > 0$ for each m . See Table 1 for

*Joshua Habiger is an Assistant Professor in the Department of Statistics at Oklahoma State University

TABLE 1

Depiction of the data in [Anderson and Habiger \(2012\)](#). Shoot biomass x_i in grams for groups $i = 1, 2, \dots, 5$ was 0.86, 1.34, 1.81, 2.37, and 3.00, respectively. Row totals are in the last column.

Bacteria m	Y_{1m}	Y_{2m}	Y_{3m}	Y_{4m}	Y_{5m}	Total (n_m)
1	0	1	1	0	5	7
2	9	2	0	0	3	14
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
778	16	10	29	18	13	81

a depiction of the data and see [Efron \(2008\)](#); [Dudoit and van der Laan \(2008\)](#); [Efron \(2010\)](#) for other, sometimes called, high dimensional (HD) data sets.

In general, multiple null hypotheses are simultaneously tested with a multiple testing procedure which, ideally, rejects as many null hypotheses as possible subject to the constraint that some global type 1 error rate is controlled at a prespecified level α . The false discovery rate (FDR) is the most frequently considered error rate in the HD setting. It is loosely defined as the expected value of the false discovery proportion (FDP), where the FDP is the proportion of erroneously rejected null hypotheses, also called false discoveries, among rejected null hypotheses, or discoveries. See [Sarkar \(2007\)](#) for other related error rates. In their seminal paper, [Benjamini and Hochberg \(1995\)](#) showed that a step-up procedure based on the [Simes \(1986\)](#) line, henceforth referred to as the BH procedure, has $\text{FDR} = \alpha a_0 \leq \alpha$ under a certain dependence structure, where a_0 is the proportion of true null hypotheses. Since then, much research has focused on developing more efficient procedures for FDR control.

One approach seeks to control the FDR at a level nearer α , as opposed to αa_0 . For example, adaptive procedures in [Benjamini and Hochberg \(2000\)](#); [Storey, Taylor and Siegmund \(2004\)](#); [Benjamini, Krieger and Yekutieli \(2006\)](#); [Gavrilov, Benjamini and Sarkar \(2009\)](#); [Liang and Nettleton \(2012\)](#) utilize an estimate of a_0 and typically have FDR that is greater than αa_0 yet still less than or equal to α . [Finner, Dickhaus and Roters \(2009\)](#) proposed nonlinear procedures that “exhaust the α ” in that, loosely speaking, their FDR converges to α under some least favorable configuration as M tends to infinity.

Another approach aims to exploit heterogeneity across hypothesis tests. [Genovese, Roeder and Wasserman \(2006\)](#); [Blanchar and Roquain \(2008\)](#); [Roquain and van de Wiel \(2009\)](#); [Peña, Habiger and Wu \(2011, 2014\)](#) propose a weighted BH-type procedure, where weights are allowed to depend on the power functions of the individual tests or prior probabilities for the states of the null hypotheses. [Cai and Sun \(2009\)](#) and [Hu, Zhao and Zhou \(2010\)](#) focused on clustered data, where test statistics are heterogeneous across clusters but homogeneous within clusters, while [Sun and McLain \(2012\)](#) considered heteroscedastic standard errors. Data in Table 1, for example, are heterogeneous because sample sizes n_1, n_2, \dots, n_M vary from test to test, with n_m being as small as 6 and as large as 911. Whatever the nature of the heterogeneity may be, recent literature suggests that it should not be ignored. Specifically, [Roeder and Wasserman \(2009\)](#) showed that weighted multiple testing procedures generally perform favorably over their unweighted counterparts,

especially when the employed weights efficiently exploit heterogeneity. In fact, [Sun and McLain \(2012\)](#) illustrated that procedures which ignore heterogeneity are not only inefficient, but can produce lists of discoveries that are of little scientific interest.

This paper combines approaches for exhausting the α and exploiting heterogeneity and shows that the resulting methodology performs favorably over competing methods. The general procedure is developed in Sections 2 - 5. Section 2 introduces a weighted adaptive multiple decision function (WAMDF) framework and a random effects model that can accommodate many types of heterogeneity, including, but not limited to, those mentioned above. Additionally, tools which facilitate implementation of the proposed WAMDF, such as weighted p -values, are developed. Section 3 derives optimal weights and explores their operating characteristics. Section 4 introduces the “asymptotically optimal” WAMDF for asymptotic FDP control and Section 5 provides a WAMDF for exact (nonasymptotic) FDR control.

Asymptotic assessment is in Sections 6. It is shown that, under a weak dependence structure, WAMDFs dominate other MDFs, asymptotically, in terms of both their FDP and power. Specifically, they are less conservative in that, asymptotically, their FDP is always larger than the FDP of their unadaptive counterpart. Further, sufficient conditions are provided under which WAMDFs provide asymptotic FDP control and exhaust the α . It is shown that these conditions are satisfied in a variety of settings, such as when optimal weights are utilized, when weights are positively correlated with optimal weights, or even in a worst case scenario where the employed weights are independent of optimal weights. As a corollary, we see that the (linear) unweighted adaptive procedure in [Storey, Taylor and Siegmund \(2004\)](#) is α -exhaustive and hence is competitive with (unweighted) nonlinear procedures in [Finner, Dickhaus and Roters \(2009\)](#). The notion of “asymptotically optimal” is formalized as well. Simulation studies in Section 7 demonstrate that WAMDFs are more powerful than competing MDFs even if the employed weights are only positively correlated with optimal weights.

In Section 8, it is demonstrated via the analysis of the data in Table 1, that WAMDFs can be useful for exploiting heterogeneity in practical applications, even if parameters for the random effects model are not precisely known. Concluding remarks are in Section 9. Technical proofs and additional detail regarding the data analysis in Section 8 are in a Supplementary section.

2. Background

2.1. Data

Let $\mathbf{Z} = (Z_m, m \in \mathcal{M})$ for $\mathcal{M} = \{1, 2, \dots, M\}$ be a random vector of test statistics with joint distribution function F and let \mathcal{F} be a model for F . The basic goal is to test null hypotheses $\mathbf{H} = (H_m, m \in \mathcal{M})$ of the form $H_m : F \in \mathcal{F}_m$, where $\mathcal{F}_m \subseteq \mathcal{F}$ is a submodel for \mathcal{F} . For short, we often denote the state of H_m

by $\theta_m = 1 - I(F \in \mathcal{F}_m)$, where $I(\cdot)$ is the indicator function, so that $\theta_m = 0(1)$ means that H_m is true(false), and denote the state of \mathbf{H} by $\boldsymbol{\theta} = (\theta_m, m \in \mathcal{M})$. Let $\mathcal{M}_0 = \{m \in \mathcal{M} : \theta_m = 0\}$ and $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_0$ index the set of true and false null hypotheses, respectively, and denote the number of true and false null hypotheses by $M_0 = |\mathcal{M}_0|$ and $M_1 = |\mathcal{M}_1|$, respectively.

To make matters concrete, we often consider the following random effects model for \mathbf{Z} . For related models see Efron et al. (2001); Genovese and Wasserman (2002); Storey (2003); Genovese, Roeder and Wasserman (2006); Sun and Cai (2007); Cai and Sun (2009); Roquain and van de Wiel (2009). In Model 1, heterogeneity across the Z_m s is attributable to prior probabilities $\mathbf{p} = (p_m, m \in \mathcal{M})$ for the states of the H_m s and parameters $\boldsymbol{\gamma} = (\gamma_m, m \in \mathcal{M})$, which we refer to as effect sizes, although each γ_m could merely index a distribution for Z_m when H_m is false.

Model 1. Let $(Z_m, \theta_m, p_m, \gamma_m), m \in \mathcal{M}$ be independent and identically distributed random vectors each with support in $\mathbb{R} \times \{0, 1\} \times [0, 1] \times \mathbb{R}^+$ and with conditional distribution functions

$$F(z_m | \theta_m, p_m, \gamma_m) = (1 - \theta_m)F_0(z_m) + \theta_m F_1(z_m | \gamma_m)$$

and

$$F(z_m | p_m, \gamma_m) = (1 - p_m)F_0(z_m) + p_m F_1(z_m | \gamma_m).$$

Assume further that $F(\gamma_m, p_m) = F(\gamma_m)F(p_m)$, $\text{Var}(\gamma_m) < \infty$ and that p_m has mean $1 - a_0 \in (0, 1)$.

Observe that Z_m has distribution function $F_0(\cdot)$ given $H_m : \theta_m = 0$ and has distribution function $F_1(\cdot | \gamma_m)$ otherwise. It is important to note that parameters $\boldsymbol{\theta}$, \mathbf{p} , and $\boldsymbol{\gamma}$ are assumed to be random variables in Model 1 primarily to facilitate asymptotic analysis, as in Genovese, Roeder and Wasserman (2006); Blanchard and Roquain (2008); Blanchard and Roquain (2009); Roquain and van de Wiel (2009); Roquain and Villers (2011). Under Model 1, focus will be on conditional distribution functions $F(\mathbf{z} | \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma}) = \prod_{m \in \mathcal{M}} F(z_m | \theta_m, p_m, \gamma_m)$ and $F(\mathbf{z} | \mathbf{p}, \boldsymbol{\gamma}) = \prod_{m \in \mathcal{M}} F(z_m | p_m, \gamma_m)$, and an expectation taken over \mathbf{Z} with respect to these distributions will be denoted $E[\cdot | \boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma}]$ and $E[\cdot | \mathbf{p}, \boldsymbol{\gamma}]$, respectively. We shall write $E[\cdot | A]$ to denote the conditional expectation of \mathbf{Z} given $(\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma}) \in A$. More generally, we write $E[\cdot]$ to denote an expectation taken with respect to some arbitrary $F \in \mathcal{F}$.

2.2. Multiple decision functions

A multiple decision function (MDF) framework is used in this paper. For similar frameworks, see Genovese and Wasserman (2004); Storey, Taylor and Siegmund (2004); Sun and Cai (2007); Peña, Habiger and Wu (2011, 2014). Specifically, each null hypothesis will be tested with the decision function $\delta_m(Z_m; t_m)$ taking values in $\{0, 1\}$ and depending on data Z_m and possibly random size threshold $t_m \in [0, 1]$, where $\delta_m = 1(0)$ means that H_m is rejected(retained). An MDF

is denoted by $\delta(\mathbf{Z}; \mathbf{t}) = [\delta_m(Z_m; t_m), m \in \mathcal{M}]$, where $\mathbf{t} = (t_m, m \in \mathcal{M})$. For example, suppose that large values of Z_m are evidence against $H_m : \theta_m = 0$ under Model 1. Then we may define

$$\delta_m(Z_m; t_m) = I(Z_m \geq F_0^{-1}(1 - t_m)). \quad (1)$$

Observe that $E[\delta_m(Z_m; t_m) | \theta_m = 0] = 1 - F_0(F_0^{-1}(1 - t_m)) = t_m$ so that t_m indeed represents the size of δ_m , hence the terminology “size threshold”. If $t_m = \alpha/M$ for each m , then $\delta(\mathbf{Z}; \mathbf{t})$ represents the well-known Bonferroni procedure.

Throughout this manuscript we assume that, for each m , $t_m \mapsto \delta_m(Z_m; t_m)$ is nondecreasing and right continuous with $\delta_m = 0(1)$ whenever $t_m = 0(1)$, almost surely, and that $t_m \mapsto E[\delta_m(Z_m; t_m)]$ is continuous and strictly increasing for $t_m \in (0, 1)$ with $E[\delta_m(Z_m; t_m)] = t_m$ whenever $m \in \mathcal{M}_0$. These assumptions are referred to as the nondecreasing-in-size (NS) assumptions and are satisfied, for example, under Model 1 for decision functions defined as in (1). For additional details and examples, see Habiger and Peña (2011); Peña, Habiger and Wu (2011); Habiger (2012).

2.3. Implementation

Computation of $\delta(\mathbf{Z}; \mathbf{t})$ is not trivial, especially when \mathbf{t} is allowed to depend on prior parameters \mathbf{p} and $\boldsymbol{\gamma}$ and data \mathbf{Z} . To simplify the problem, we break \mathbf{t} down into the product of a positive valued weight vector $\mathbf{w} = (w_m, m \in \mathcal{M})$ satisfying $\bar{w} = M^{-1} \sum_{m \in \mathcal{M}} w_m = 1$ and an overall or average threshold t . That is, we write $\mathbf{t} = t\mathbf{w}$. The basic methodology is then operationally implemented in the following two steps:

1. specify weights \mathbf{w} and
2. collect data $\mathbf{Z} = \mathbf{z}$, find t , and compute $\delta(\mathbf{z}; t\mathbf{w})$.

As in Genovese, Roeder and Wasserman (2006), weights are allowed to depend functionally on prior parameters \mathbf{p} and $\boldsymbol{\gamma}$ in Model 1 in the first step, while the overall threshold t will be allowed to depend functionally on data \mathbf{Z} and \mathbf{w} in the second step. The Weight Selection Procedure for the first step is formally defined in in Section 3 and the Threshold Selection Procedure for the second step is defined in Section 4.

However, before proceeding we develop some foundational tools which facilitate the implementation of the second step. Define the (unweighted) p -value statistic corresponding to δ_m by

$$P_m = \inf\{t_m \in [0, 1] : \delta_m(Z_m; t_m) = 1\}. \quad (2)$$

This definition, also considered in Habiger and Peña (2011); Peña, Habiger and Wu (2011), has the usual interpretation that P_m is the smallest size t_m allowing for H_m to be rejected and also ensures that

$$\delta_m(Z_m; t_m) = I(P_m \leq t_m)$$

almost surely under the NS assumptions. For example, it can be verified that the p -value statistic corresponding to (1) is $P_m = 1 - F_0(Z_m)$ and that $I(Z_m \leq F_0^{-1}(1 - t_m)) = I(P_m \leq t_m)$ almost surely. See Habiger (2012); Habiger and Peña (2014) for more details or for derivations of more complex p -values, such as the p -value for the local FDR statistic in Efron et al. (2001); Sun and Cai (2007) or for the optimal discovery procedure in Storey (2007).

Now define weighted p -value by

$$Q_m = \inf\{t : \delta_m(Z_m; tw_m) = 1\}. \quad (3)$$

Observe that for w_m fixed and writing $t_m = tw_m$,

$$P_m = \inf\{tw_m : \delta_m(Z_m; tw_m) = 1\} = w_m \inf\{t : \delta_m(Z_m; tw_m) = 1\} = w_m Q_m$$

almost surely. Thus, a weighted p -value can be computed by $Q_m = P_m/w_m$. Hence, we have established the following almost surely equivalent expressions for a decision function under the NS assumptions:

$$\delta_m(Z_m; t_m) = \delta_m(Z_m; tw_m) = I(P_m \leq tw_m) = I(Q_m \leq t). \quad (4)$$

3. Optimal weights

We first focus on optimal weights for Model 1 which assume that the overall threshold t in $\delta(\mathbf{Z}; tw)$ is fixed. These weights are directly implementable in a variety of MDFs, such as the weighted Bonferroni-type procedure in Spjøtvoll (1972), where $t = \alpha/M$. They are also useful for our proposed WAMDF because, even though it utilizes a data dependent overall threshold and weights must be specified before data are collected, the threshold can be well approximated for large M using only \mathbf{p} and $\boldsymbol{\gamma}$. We refer to the first set of weights as optimal fixed- t weights and refer to the latter set of weights based on the approximated threshold as “asymptotically optimal” for reasons that will be formally justified in Theorem 8 (see Section 6).

3.1. Optimal fixed- t weights

For the moment, we focus on $\delta(\mathbf{Z}; t)$ rather than $\delta(\mathbf{Z}; tw)$. Recall that these expressions are equivalent due to (10.3) and the constraint that t is fixed with $\bar{w} = 1$ is equivalent to the constraint that $\bar{t} = t$, where $\bar{t} = M^{-1} \sum_{m \in \mathcal{M}} t_m$. Now, because weights are allowed to depend on \mathbf{p} and $\boldsymbol{\gamma}$ under Model 1, we focus on

$$G_m(t_m) \equiv E[\delta_m(Z_m; t_m) | \mathbf{p}, \boldsymbol{\gamma}] = (1 - p_m)t_m + p_m \pi_{\gamma_m}(t_m), \quad (5)$$

where $\pi_{\gamma_m}(t_m) \equiv E[\delta_m(Z_m; t_m) | \theta_m = 1, \gamma_m]$ is the power function for δ_m . As in Genovese, Roeder and Wasserman (2006); Roquain and van de Wiel (2009); Peña, Habiger and Wu (2011), we assume power functions (as a function of t_m) are concave.

- (A1) For each $m \in \mathcal{M}$, $t_m \mapsto \pi_{\gamma_m}(t_m)$ is concave and twice differentiable for $t_m \in (0, 1)$ with $\lim_{t_m \uparrow 1} \pi'_{\gamma_m}(t_m) = 0$ and $\lim_{t_m \downarrow 0} \pi'_{\gamma_m}(t_m) = \infty$ almost surely, where $\pi'_{\gamma_m}(t_m)$ is the derivative of $\pi_{\gamma_m}(t_m)$ with respect to t_m .

This concavity condition is satisfied, for example, under monotone likelihood ratio considerations (Peña, Habiger and Wu, 2011) and under the generalized monotone likelihood ratio (GMLR) condition in Cao, Sun and Kosorok (2013). The GMLR condition states that $g_{1m}(t_m)/g_{0m}(t_m)$ is monotonically decreasing in t_m , where $g_{im}(t_m)$ is the derivative of $G_{im}(t_m) \equiv E[\delta_m(Z_m; t_m) | \theta_m = i]$ with respect to t_m for $i = 0, 1$. In our notation, $g_{1m}(t_m) = \pi'_{\gamma_m}(t_m)$ and, under the NS conditions, $g_{0m}(t_m) = 1$. Hence, the GMLR condition stipulates that $\pi'_{\gamma_m}(t_m)$ is monotonically decreases, i.e. that $\pi_{\gamma_m}(t_m)$ is concave.

Given \mathbf{p} , γ and t , the goal is to maximize the expected number of correctly rejected null hypotheses

$$\pi(\mathbf{t}, \mathbf{p}, \gamma) \equiv E \left[\sum_{m \in \mathcal{M}} \theta_m \delta_m(Z_m; t_m) \middle| \gamma, \mathbf{p} \right] = \sum_{m \in \mathcal{M}} p_m \pi_{\gamma_m}(t_m), \quad (6)$$

subject to the constraint that $\bar{t} = t$. Theorem 1 describes the form of the solution and states that it exists and is unique.

Theorem 1. *Suppose that (A1) is satisfied and fix $t \in (0, 1)$. Then under Model 1 the maximum of $\pi(\mathbf{t}, \mathbf{p}, \gamma)$ with respect to \mathbf{t} subject to constraint $\bar{t} = t$ exists, is unique, and satisfies*

$$\pi'_{\gamma_m}(t_m) = k/p_m \quad (7)$$

for every $m \in \mathcal{M}$ and some $k > 0$ almost surely.

The optimal fixed- t thresholds described in Theorem 1, and their corresponding weights, can be numerically found as follows. First, for a fixed value of k , denote the solution to equation (7) in terms of t_m by $t_m(k/p_m, \gamma_m)$ and denote the vector of solutions by $\mathbf{t}(k, \mathbf{p}, \gamma) = [t_m(k/p_m, \gamma_m), m \in \mathcal{M}]$. In Example 1 below a closed form expression for $t_m(k/p_m, \gamma_m)$ exists, but in general we may employ any single root finding algorithm to compute each $t_m(k/p_m, \gamma_m)$ because $\pi'_{\gamma_m}(t_m)$ is continuous and monotone in t_m by (A1). Then to find the optimal fixed- t threshold vector, first find the unique k^* satisfying $\bar{t}_M(k^*, \mathbf{p}, \gamma) = t$, where $\bar{t}_M(k, \mathbf{p}, \gamma) = M^{-1} \sum_{m \in \mathcal{M}} t_m(k/p_m, \gamma_m)$, and then compute $\mathbf{t}(k^*, \mathbf{p}, \gamma)$. Because $t_m = tw_m$ in our setup and because $\bar{t}_M(k^*, \mathbf{p}, \gamma) = t$, each optimal fixed- t weight is recovered via

$$w_m(k^*, \mathbf{p}, \gamma) = \frac{t_m(k^*/p_m, \gamma_m)}{\bar{t}_M(k^*, \mathbf{p}, \gamma)}. \quad (8)$$

We shall sometimes denote $w_m(k^*, \mathbf{p}, \gamma)$ by w_m^* for short. The vector of optimal fixed- t weights is denoted by $\mathbf{w}(k^*, \mathbf{p}, \gamma) = [w_m(k^*, \mathbf{p}, \gamma), m \in \mathcal{M}]$ or by $\mathbf{w}^* = (w_m^*, m \in \mathcal{M})$ for short.

To better understand how the solution is found and related to the values of p_m , γ_m and t consider the following example.

Example 1. Suppose $Z_m|\gamma_m, \theta_m \sim N(\theta_m\gamma_m, 1)$ for $\gamma_m > 0$, where $N(a, b)$ represents a Normal distribution with mean a and variance b , and consider testing $H_m : \theta_m = 0$. Denote the standard normal cumulative distribution function and density function by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively, and define $\delta_m(Z_m; t_m) = I(Z_m \geq \Phi^{-1}(1 - t_m))$. The power function is

$$\pi_{\gamma_m}(t_m) = 1 - \Phi(\Phi^{-1}(1 - t_m) - \gamma_m)$$

and has derivative

$$\pi'_{\gamma_m}(t_m) = \frac{\phi(\Phi^{-1}(1 - t_m) - \gamma_m)}{\phi(\Phi^{-1}(1 - t_m))}.$$

Setting the derivative equal to k/p_m and solving yields

$$t_m(k/p_m, \gamma_m) = 1 - \Phi(0.5\gamma_m + \log(k/p_m)/\gamma_m). \quad (9)$$

The optimal fixed- t threshold vector is computed by $\mathbf{t}(k^*, \mathbf{p}, \boldsymbol{\gamma})$, where k^* satisfies $\bar{t}_M(k^*, \mathbf{p}, \boldsymbol{\gamma}) = t$, and the optimal fixed- t weights are computed as in (8).

First, observe in expression (9) that $t_i(k/p_i, \gamma_i) = t_j(k/p_j, \gamma_j)$ if $\gamma_i = \gamma_j$ and $p_i = p_j$ regardless of k , and consequently, the optimal fixed- t weight vector is $\mathbf{1}$ for any t when data are homogeneous. On the other hand, we see that $t_m(k/p_m, \gamma_m)$ is increasing in p_m and hence

$$w_m(k^*, \mathbf{p}, \boldsymbol{\gamma}) = M \frac{t_m(k^*/p_m, \gamma_m)}{t_m(k^*/p_m, \gamma_m) + \sum_{j \neq m} t_j(k^*/p_j, \gamma_j)}$$

is increasing in p_m . Hence, if we increase the likelihood that H_m is false then the corresponding optimal weight increases.

The relationship between $w_m(k^*, \mathbf{p}, \boldsymbol{\gamma})$ and γ_m is more complex. To illustrate, consider testing $M = 2$ null hypotheses and suppose $\gamma_1 = 1.5$, $\gamma_2 = 2.5$, and $p_1 = p_2 = 0.5$. In Figure 1, observe that for $t = 0.01$, $\bar{t}_M(k^*, \mathbf{p}, \boldsymbol{\gamma}) = 0.01$ when $k^* = 6.1$, which gives $t_1(k^*/p_1, \gamma_1) = 0.017$, $t_2(k^*/p_2, \gamma_2) = 0.003$, $w_1^* = 0.003/0.01 = 0.3$ and $w_2^* = 0.017/0.01 = 1.7$. Note that because $p_1 = p_2$, the slopes of the power functions evaluated at 0.003 and 0.017, respectively, are equal; see equation (7). Now consider fixed threshold $t = 0.05$. Here $k^* = 1.7$ and the slopes of the power functions evaluated at the solutions are again equal, but now $w_1^* = 0.059/0.05 = 1.18$ and $w_2^* = 0.041/0.05 = 0.82$. That is, when $t = 0.01$, the hypothesis with the larger effect size is given more weight but when $t = 0.05$ it is given less weight. For a more detailed discussion on this phenomenon see Peña, Habiger and Wu (2011) or Section 8 of the current manuscript. The important point is that the optimal fixed- t weights depend on the choice of t , in addition to the parameters \mathbf{p} and $\boldsymbol{\gamma}$.

3.2. Asymptotically optimal weights

In Section 4, the overall threshold t is chosen using an estimator of the FDP which depends functionally on data \mathbf{Z} ; see expressions (10) and (11). Hence,

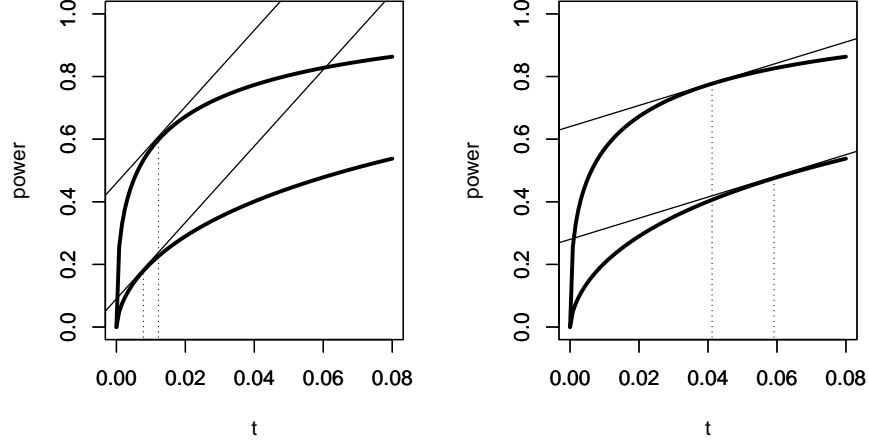


FIG 1. A depiction of the optimal thresholds for $M = 2$ hypotheses tests when power functions vary under constraint $0.5(t_1 + t_2) = 0.01$ (left) and $0.5(t_1 + t_2) = 0.05$ (right).

optimal fixed- t weights, which are not allowed to functionally depend on \mathbf{Z} , are not readily implementable. To solve this problem, we approximate the FDP estimator using only \mathbf{p} and $\boldsymbol{\gamma}$ and use the resulting “approximator” to approximate the data dependent threshold. Then, the optimal fixed- t weights for the approximated threshold are computed. The approximator essentially plugs $G_m(t_m(k/p_m, \gamma_m))$ in for each $\delta_m(Z_m; t_m(k/p_m, \gamma_m))$ in (10) and (11). Formally, denote $\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma})) = M^{-1} \sum_{m \in \mathcal{M}} G_m(t_m(k/p_m, \gamma_m))$ and define the FDP approximator by

$$\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma})) = \frac{1 - \bar{G}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma}))}{1 - \bar{t}_M(k, \mathbf{p}, \boldsymbol{\gamma})} \frac{\bar{t}_M(k, \mathbf{p}, \boldsymbol{\gamma})}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma}))}.$$

Now, the asymptotically optimal weights for the proposed WAMDF are computed as follows for $0 < \alpha \leq 1 - p_{(M)}$, where $p_{(M)} \equiv \max\{\mathbf{p}\}$.

Weight Selection Procedure. For $0 < \alpha \leq 1 - p_{(M)}$,

- a. get $k_M^* = \inf \left\{ k : \widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma})) = \alpha \right\}$, and
- b. for each $m \in \mathcal{M}$, compute $w_m^* = w_m(k_M^*, \mathbf{p}, \boldsymbol{\gamma})$ as in (8).

In Theorem 2 below, we see that the restriction $0 < \alpha \leq 1 - p_{(M)}$ ensures that a solution to $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \boldsymbol{\gamma})) = \alpha$ exists. In practice, this restriction amounts to choosing α and \mathbf{p} so that $0 < \alpha \leq 1 - p_m$ for each m . That is, the prior probability that the null hypothesis is true should be at least α . To see why this condition is reasonable, suppose that $1 - p_m < \alpha$ for each m .

Then we need not consider a weighting scheme or even collect data in the first place because even if all M null hypotheses are rejected, the model stipulates that the expected proportion of false discoveries among the M discoveries is $M^{-1} \sum_{m \in \mathcal{M}} 1 - p_m < \alpha$. Hence, if the conditions of Theorem 2 are not satisfied, we should suspect that the posited model is poor and consider a new model.

Theorem 2. *Under (A1) and Model 1, k_M^* exists almost surely for $0 < \alpha \leq 1 - p_{(M)}$.*

It should be noted that $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma))$ need not be monotone in k and hence multiple solutions to $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) = \alpha$ could exist. We choose the smallest k above because this solution has the largest overall threshold $\bar{t}_M(k, \mathbf{p}, \gamma)$ (recall $\bar{t}_M(k, \mathbf{p}, \gamma)$ increases as k decreases) and hence is consistent with our data dependent threshold selection procedure in the next section. Also note that $\bar{t}_M(k_M^*, \mathbf{p}, \gamma) = t$ for some $t \in (0, 1)$, i.e. these weights could be viewed as optimal fixed- t weights. The main difference is that the constraint $\widetilde{FDP}_M(\mathbf{t}(k_M^*, \mathbf{p}, \gamma)) = \alpha$ is specified for the asymptotically optimal weights, as opposed to the constraint $\bar{t}_M(k_M^*, \mathbf{p}, \gamma) = t$ in Theorem 1. We use the notation k_M^* , as opposed to k^* , to indicate that we are dealing specifically with asymptotically optimal weights and to facilitate asymptotic analysis later.

4. The procedure

Now we are now in position to formally define the proposed adaptive threshold, which, when used in conjunction with asymptotically optimal weights in $\delta(\mathbf{Z}; t\mathbf{w})$, yields the proposed asymptotically optimal WAMDF.

4.1. Threshold selection

For the moment, let \mathbf{w} be any fixed vector of positive weights satisfying $\bar{w} = 1$. For brevity, we sometimes suppress the Z_m in each δ_m and write $\delta_m(t\mathbf{w}_m)$ and denote $\delta(\mathbf{Z}; t\mathbf{w})$ by $\delta(t\mathbf{w})$. Further, denote the number of rejected null hypotheses, or discoveries, at $t\mathbf{w}$ by $R(t\mathbf{w}) = \sum_{m \in \mathcal{M}} \delta_m(t\mathbf{w}_m)$.

We make use of an “adaptive” estimator of the FDP, i.e. it utilizes an estimator of M_0 defined by

$$\hat{M}_0(\lambda\mathbf{w}) = \frac{M - R(\lambda\mathbf{w}) + 1}{1 - \lambda} \quad (10)$$

for some fixed tuning parameter $\lambda \in (0, 1)$. This estimator is essentially the weighted version of the estimator in Storey (2002) defined by $\hat{M}_0(\lambda\mathbf{1}) = [M - R(\lambda\mathbf{1})]/[1 - \lambda]$. For earlier work on the estimation of M_0 see Schweder and Spjøtvoll (1982). The idea, in the unweighted setting, is that for $m \in \mathcal{M}_1$, $E[\delta_m(\lambda)] \leq 1$ but the inequality is relatively sharp if all tests have reasonable power and λ is chosen sufficiently large. Hence

$$E[M - R(\lambda\mathbf{1})] = \sum_{m \in \mathcal{M}} E[1 - \delta_m(\lambda)] \geq \sum_{m \in \mathcal{M}_0} E[1 - \delta_m(\lambda)] = (1 - \lambda)M_0$$

and $E[\hat{M}_0(\lambda \mathbf{1})] \geq M_0$. That is, \hat{M}_0 is positively biased but the bias is minor for suitably chosen λ and reasonably powerful tests. Similar intuition applies for $\hat{M}_0(\lambda \mathbf{w})$. As in [Storey, Taylor and Siegmund \(2004\)](#), we add 1 to the numerator in expression (10) to ensure that $\hat{M}_0(\lambda \mathbf{w}) > 0$ for finite sample results. For a discussion on weighted vs. unweighted estimation of M_0 , see Section 9.

The adaptive FDP estimator is defined by

$$\widehat{FDP}^\lambda(t\mathbf{w}) = \frac{\hat{M}_0(\lambda \mathbf{w})t}{\max\{R(t\mathbf{w}), 1\}}. \quad (11)$$

The adaptive threshold, which essentially chooses t as large as possible subject to the constraint that the estimate of the FDP is less than or equal to α , is defined by

$$\hat{t}_\alpha^\lambda = \sup\{0 \leq t \leq u : \widehat{FDP}^\lambda(t\mathbf{w}) \leq \alpha\}. \quad (12)$$

We assume that u , the upper bound for \hat{t}_α^λ , and the tuning parameter λ satisfy condition (A2), given by

$$(A2) \quad \lambda \leq u \leq 1/w_{(M)},$$

where $w_{(M)} \equiv \max\{\mathbf{w}\}$. This ensures that $\hat{t}_\alpha^\lambda w_m \leq 1$ and $\lambda w_m \leq 1$ for every m . It should be noted that for $\mathbf{w} = \mathbf{1}$ and $u = \lambda$ (which implies $\hat{t}_\alpha^\lambda \leq \lambda$), we recover the unweighted adaptive MDF for finite FDR control in [Storey, Taylor and Siegmund \(2004\)](#).

In practice, \hat{t}_α^λ can be difficult to compute. Alternatively, we may apply the original BH procedure to the weighted p -values at level $\alpha/\hat{M}_0(\lambda \mathbf{w})$. Note that due to (4) we can also use weighted p -values to estimate M_0 via

$$\hat{M}_0(\lambda \mathbf{w}) = \frac{M - \sum_{m \in \mathcal{M}} I(Q_m \leq \lambda) + 1}{1 - \lambda}.$$

Formally, this threshold selection procedure can be implemented as follows.

Threshold Selection Procedure. Fix λ and u satisfying (A2). Then

- a. compute weighted p -values as in (3) and get ordered weighted p -values $Q_{(1)} \leq Q_{(2)} \leq \dots \leq Q_{(M)}$.
- b. If $Q_{(m)} > \alpha m / \hat{M}_0(\lambda \mathbf{w})$ for each m , set $j = 0$, otherwise take

$$j = \max\left\{m \in \mathcal{M} : Q_{(m)} \leq \alpha / \hat{M}_0(\lambda \mathbf{w})\right\}.$$

- c. Get $\hat{t}_\alpha^{\lambda*} = \min\{j\alpha / \hat{M}_0(\lambda \mathbf{w}), u\}$ and compute $\delta_m(Z_m; \hat{t}_\alpha^{\lambda*} w_m) = I(Q_m \leq \hat{t}_\alpha^{\lambda*})$ for each m .

The WAMDF implemented above is equivalent to $\delta(\mathbf{Z}; \hat{t}_\alpha^\lambda \mathbf{w})$ in that

$$\delta_m(Z_m; \hat{t}_\alpha^\lambda w_m) = I(Q_m \leq \hat{t}_\alpha^\lambda) = I(Q_m \leq \hat{t}_\alpha^{\lambda*}) \quad (13)$$

almost surely for each m , i.e. both procedures reject the same set of null hypotheses. The first equality in (13) follows from (4) and the last equality in (13) is a consequence of Lemma 2 in [Storey, Taylor and Siegmund \(2004\)](#).

TABLE 2

A portion of the parameters, data, weights, p -values, and weighted p -values in columns 1 - 5, respectively. Each row is sorted in ascending order according to Q_1, Q_2, \dots, Q_M .

θ_m	γ_m	w_m^*	Z_m	P_m	Q_m	$0.05m/\tilde{M}_0$
1	3	0.74	3.14	0.001	0.001	0.006
1	2	1.26	2.55	0.005	0.005	0.012
1	3	0.74	2.56	0.005	0.006	0.018
1	2	1.26	1.47	0.070	0.062	0.024
0	2	1.74	1.17	0.121	0.106	0.030
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
0	3	0.74	-0.60	0.724	0.844	0.061

4.2. The asymptotically optimal WAMDF

The asymptotically optimal WAMDF is formally defined as $\delta(\mathbf{Z}; \hat{t}_\alpha^\lambda \mathbf{w}^*)$ for $0 < \alpha \leq 1 - p_{(M)}$ and $\lambda = \bar{t}_M(k_M^*, \mathbf{p}, \gamma)$, and where k_M^* and \mathbf{w}^* are defined as in the Weight Selection Procedure. This particular choice of λ is not necessary for ensuring asymptotic FDP or FDR control but, as we will see, ensures that the employed weights are indeed “asymptotically optimal” (see Theorem 8) and additionally that (A2) is satisfied if we take $u = 1/w_{(M)}$. The asymptotically optimal WAMDF is implemented as follows.

Asymptotically Optimal wamdf. To implement the asymptotically optimal WAMDF

1. specify (\mathbf{p}, γ) , $\alpha \in (0, 1 - p_{(M)}]$ and compute \mathbf{w}^* using the Weight Selection Procedure,
2. choose $\lambda = \bar{t}_M(k_M^*, \mathbf{p}, \gamma)$ and u satisfying (A2), collect data $\mathbf{Z} = \mathbf{z}$ and compute $\delta(\mathbf{z}; \hat{t}_\alpha^\lambda \mathbf{w}^*)$ using the Threshold Selection Procedure.

To illustrate, consider testing $M = 10$ null hypotheses under the setting outlined in Example 1, with $p_m = 0.5$ for $m = 1, 2, \dots, 10$, $\gamma_m = 2$ for $m = 1, 2, \dots, 5$, $\gamma_m = 3$ for $m = 6, 7, \dots, 10$, and take $\alpha = 0.05$. Recall the goal is to test $H_m : \theta_m = 0$ and that decision functions are of the form $\delta_m(Z_m; t_m) = I(Z_m \geq \Phi^{-1}(1 - t_m))$. Further, p -values are computed by $P_m = 1 - \Phi(Z_m)$ and recall that weighted p -values can be computed by $Q_m = P_m/w_m$. See Table 2 for summaries of parameters, weights, simulated data, p -values, and weighted p -values. As before, the Weight Selection Procedure in step 1 is broken down into 2 sub-steps and the Threshold Selection Procedure in step 2 is split into three sub-steps. Now, to test these null hypotheses we

- 1a. specify γ (see column 2 of Table 2), \mathbf{p} and α and find $k_M^* = 2.52$.
- 1b. Compute asymptotically optimal weights $w_m^* = w_m(k_M^*, \mathbf{p}, \gamma)$ as in (8). See column 3 in Table 2.
- 2a. Take $\lambda = \bar{t}_M(k_M^*, \mathbf{p}, \gamma) = 0.028$ and $u = 1/1.26 = 0.79$. Collect data $\mathbf{Z} = \mathbf{z}$, compute p -values and weighted p -values, and order weighted p -values from smallest to largest (see columns 4 - 6 in Table 2).

- 2b. Observe that $Q_{(m)} \leq \alpha m / \hat{M}_0(\lambda \mathbf{w}^*)$ for $m = 3$ but not for $m = 4, 5, \dots, 10$ and hence $\alpha j / \hat{M}_0(\lambda \mathbf{w}^*) = 0.05 \frac{3}{8.23} = 0.013$.
- 2c. Compute $\hat{t}_\alpha^{\lambda*} = \min\{0.013, 0.79\} = 0.013$ and reject null hypotheses with weighted p -values 0.001, 0.005 and 0.006 because their Q_{ms} are less than 0.013.

5. Finite FDR control

Next an upper bound for the FDR is given for any arbitrary set of fixed or realized weights satisfying $w_m > 0$ for each m and $\bar{w} = 1$. The bound is computed under the following dependence structure for \mathbf{Z} :

- (A3) $(Z_m, m \in \mathcal{M}_0)$ are mutually independent and independent of the collection $(Z_m, m \in \mathcal{M}_1)$.

This structure was considered in [Benjamini and Hochberg \(1995\)](#) (and relaxed in [Benjamini and Yekutieli \(2001\)](#)) to prove FDR control for the original unweighted unadaptive BH procedure. It was also used in the proof of FDR control for the weighted unadaptive BH procedures in [Genovese, Roeder and Wasserman \(2006\)](#); [Peña, Habiger and Wu \(2011, 2014\)](#) and for the unweighted adaptive BH procedure in [Storey, Taylor and Siegmund \(2004\)](#). Note that (A3) is satisfied under Model 1 conditionally upon $\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma}$, i.e. for $F(\mathbf{z}|\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma})$, but is not limited to this setting.

To define the FDR, let $V(t\mathbf{w}) = \sum_{m \in \mathcal{M}_0} \delta_m(tw_m)$ denote the number of erroneously rejected null hypotheses (false discoveries) and recall that $R(t\mathbf{w}) = \sum_{m \in \mathcal{M}} \delta_m(t\mathbf{w})$ is the number of rejected null hypotheses, or discoveries, for threshold vector $t\mathbf{w}$. Define the FDP at $t\mathbf{w}$ by

$$FDP(t\mathbf{w}) = \frac{V(t\mathbf{w})}{\max\{R(t\mathbf{w}), 1\}}. \quad (14)$$

The FDR at $t\mathbf{w}$ is defined by $FDR(t\mathbf{w}) = E[FDP(t\mathbf{w})]$, where the expectation is taken over \mathbf{Z} with respect to an arbitrary $F \in \mathcal{F}$, which may include, but is not limited to, $F(\mathbf{z}|\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\gamma})$ in Model 1.

The bound is presented in Lemma 1 below for any distribution function F satisfying (A3). The focus is on the setting when $M_0 \geq 1$ because the FDR is trivially 0 if $M_0 = 0$. As in [Storey, Taylor and Siegmund \(2004\)](#), we force $\hat{t}_\alpha^\lambda \leq \lambda$ by taking $u = \lambda$ in (12) to facilitate the use of the Optimal Stopping Theorem in the proof.

Lemma 1. *Suppose $M_0 \geq 1$ and that (A2) and (A3) are satisfied. Then for \hat{t}_α^λ defined as in (12) with $u = \lambda$,*

$$FDR(\hat{t}_\alpha^\lambda \mathbf{w}) \leq \alpha \bar{w}_0 \frac{1 - \lambda}{1 - \lambda \bar{w}_0} [1 - (\lambda \bar{w}_0)^{M_0}] \leq \alpha \bar{w}_0 \frac{1 - \lambda}{1 - \lambda \bar{w}_0}, \quad (15)$$

where $\bar{w}_0 = M_0^{-1} \sum_{m \in \mathcal{M}_0} w_m$ is the mean of the weights corresponding to true null hypotheses.

Observe that $1 - (\lambda \bar{w}_0)^{M_0} \leq 1$ due to (A2). Further, if $\mathbf{w} = \mathbf{1}$ then $\bar{w}_0 = 1$ and we recover Theorem 3 in Storey, Taylor and Siegmund (2004) as a corollary.

Of course, if $\mathbf{w} \neq \mathbf{1}$, the bound in Lemma 1 is not immediately applicable because M_0 and consequently \bar{w}_0 is unobservable. One solution is to utilize an upper bound for \bar{w}_0 and adjust the “ α ” at which the procedure is applied. The necessary α -adjustment is presented in Theorem 3.

Theorem 3. *Define*

$$\alpha^* = \alpha \frac{1}{w_{(M)}} \frac{1 - \lambda w_{(M)}}{1 - \lambda}.$$

Then under the conditions of Lemma 1, $FDR(\hat{t}_{\alpha^}^\lambda \mathbf{w}) \leq \alpha$.*

In the next section, we see that \bar{w}_0 is typically less than or equal to 1, asymptotically, so that this α adjustment is not needed for large M .

6. Asymptotic results

We first focus on an asymptotic assessment of WAMDFs for arbitrary weights. In particular, we show that they dominate their unadaptive counterparts in that they always reject more null hypotheses. This generalizes the results in Storey, Taylor and Siegmund (2004), where \mathbf{w} was equal to $\mathbf{1}$. Then, sufficient conditions for the weights are provided under which WAMDFs control the FDP and are α -exhaustive. These results are used in the asymptotic analysis of the asymptotically optimal WAMDF.

To facilitate asymptotic analysis, denote weight vectors of length M by \mathbf{w}_M and the m th element of \mathbf{w}_M by $w_{m,M}$. Further, denote the mean of the weights from true null hypotheses by $\bar{w}_{0,M}$. Denote the adaptive FDP estimator in (11) by $\widehat{FDP}_M^\lambda(t\mathbf{w}_M)$ and the FDP in (14) by $FDP_M(t\mathbf{w}_M)$. We will also consider an unadaptive FDP estimator, which uses M in the place of an estimate of M_0 , defined by

$$\widehat{FDP}_M^0(t\mathbf{w}_M) = \frac{Mt}{\max\{R(t\mathbf{w}_M), 1\}}.$$

When necessary, we also denote the tuning parameter in (10) by λ_M because, as in the asymptotically optimal WAMDF where $\lambda_M = \bar{t}_M(k_M^*, \mathbf{p}, \gamma)$, it may depend on M .

Recall that we assumed $\lambda \leq u \leq 1/w_{(M)}$ in (A2) to ensure that every individual threshold was bounded above by 1. In our asymptotic analysis, we redefine (A2) as follows:

(A2) $\lambda_M \rightarrow \lambda \leq u = 1/k$ almost surely, where k satisfies $\lim_{M \rightarrow \infty} w_{(M)} \leq k$ almost surely.

We will see that (A2) is satisfied, for example, under Model 1 and (A1) for the asymptotically optimal WAMDF. Now, the adaptive threshold is defined as in (12) and is denoted $\hat{t}_{\alpha,M}^\lambda$. The unadaptive threshold is defined by

$$\hat{t}_{\alpha,M}^0 = \sup\{0 \leq t \leq u : \widehat{FDP}_M^0(t\mathbf{w}_M) \leq \alpha\}.$$

6.1. Arbitrary weights

Convergence criteria considered here are similar to criteria in [Storey, Taylor and Siegmund \(2004\)](#); [Genovese, Roeder and Wasserman \(2006\)](#) and allow for weak dependence structures. See [Billingsley \(1999\)](#), [Storey \(2003\)](#) or see Theorem 7 for examples. Assume for any fixed $t \in (0, u]$ that

- (A4) $R(t\mathbf{w}_M)/M \rightarrow G(t)$ almost surely,
- (A5) $V(t\mathbf{w}_M)/M \rightarrow a_0\mu_0 t$ almost surely, for $0 < \mu_0 < \infty$ and $0 < a_0 < 1$, where $\bar{w}_{0,M} \rightarrow \mu_0$ and $M_0/M \rightarrow a_0$, and
- (A6) $t/G(t)$ is strictly increasing and continuous over $(0, u)$ with

$$\lim_{t \downarrow 0} \frac{t}{G(t)} = 0 \quad \text{and} \quad \lim_{t \uparrow u} \frac{u}{G(u)} \leq 1.$$

Here μ_0 is the asymptotic mean of the weights corresponding to true null hypotheses and a_0 is the asymptotic proportion of true null hypotheses. The last condition is natural as it will ensure that, asymptotically, the FDP is continuous and increasing in t and takes on value 0, thereby ensuring that it can be controlled. Note that writing $R(t\mathbf{w}_M)/M = \sum_{m \in \mathcal{M}} I(Q_m \leq t)/M$ via (4), we see that (A4) corresponds to the assumption that the empirical process of the weighted p -values converges pointwise to $G(t)$ almost surely.

Asymptotic analysis will focus on comparing random thresholds $\hat{t}_{\alpha, M}^\lambda$ and $\hat{t}_{\alpha, M}^0$ to their corresponding asymptotic (nonrandom) thresholds, which are based on the limits of the unadaptive and adaptive FDP estimators. Denote the pointwise limits of the unadaptive FDP estimator, the adaptive FDP estimator and the FDP by

$$FDP_\infty^0(t) = \frac{t}{G(t)}, \quad FDP_\infty^\lambda(t) = \frac{1 - G(\lambda)}{1 - \lambda} \frac{t}{G(t)} \quad \text{and} \quad FDP_\infty(t) = \frac{a_0\mu_0 t}{G(t)},$$

respectively (see Lemma S1 in the Supplement for verification and details). Define asymptotic unadaptive and asymptotic adaptive thresholds by

$$t_{\alpha, \infty}^0 = \sup\{0 \leq t \leq u : FDP_\infty^0(t) \leq \alpha\}$$

and

$$t_{\alpha, \infty}^\lambda = \sup\{0 \leq t \leq u : FDP_\infty^\lambda(t) \leq \alpha\},$$

respectively.

In Theorem 4 we see that both the unadaptive and adaptive thresholds converge to their asymptotic (nonrandom) counterparts, with the asymptotic adaptive threshold being larger than the asymptotic unadaptive threshold. By virtue of $E[\delta_m(t\mathbf{w}_m)]$ being strictly increasing in t for each m , it follows that the adaptive procedure will lead to a higher proportion of rejected null hypotheses, asymptotically.

Theorem 4. Fix $\alpha \in (0, 1)$. Then under (A2) and (A4) - (A6)

$$\lim_{M \rightarrow \infty} \hat{t}_{\alpha, M}^0 = t_{\alpha, \infty}^0 \leq \lim_{M \rightarrow \infty} \hat{t}_{\alpha, M}^\lambda = t_{\alpha, \infty}^\lambda \quad (16)$$

almost surely.

Before focusing on the FDP it is useful to formally describe the notion of an α -exhaustive MDF. Loosely speaking, [Finner, Dickhaus and Roters \(2009\)](#) referred to an unweighted multiple decision function, say $\delta(\hat{t}_{\alpha,M}^* \mathbf{1}_M)$, as “asymptotically optimal” (we will use the terminology α -exhaustive) if $FDR(\hat{t}_{\alpha,M}^* \mathbf{1}_M) \rightarrow \alpha$ under some least favorable distribution. A Dirac Uniform (DU) distribution, which we shall define as a distribution that satisfies

$$E[\delta_m(t)] = \begin{cases} t & \text{if } m \in \mathcal{M}_0 \\ 1 & \text{if } m \in \mathcal{M}_1 \end{cases}$$

for every $t \in [0, 1]$ and $m \in \mathcal{M}$, was shown to often be least favorable for the FDR in that among all F s that satisfy $E[\delta_m(t)] = t$ for every $t \in [0, 1]$ when $m \in \mathcal{M}_0$ and dependency structure (A3), $FDR(\hat{t}_{\alpha,M}^* \mathbf{1}_M)$ is the largest under a DU distribution. Observe that if (A4) - (A5) are satisfied, then $G(t) = a_0 \mu_0 t + (1 - a_0)$ under a DU distribution for $t \leq u$. Denote this particular $G(t)$ by $G^{DU}(t)$.

To study asymptotic FDP of the WAMDF we consider

$$\lim_{M \rightarrow \infty} FDP_M(\hat{t}_{\alpha,M}^0 \mathbf{w}_M) \leq \lim_{M \rightarrow \infty} FDP_M(\hat{t}_{\alpha,M}^\lambda \mathbf{w}_M) \leq \alpha \quad (17)$$

and the following three claims regarding the inequalities:

- (C1) the first inequality in (17) is satisfied almost surely,
- (C2) the second inequality in (17) is satisfied almost surely, and
- (C3) the second inequality in (17) is an equality almost surely under a DU distribution.

Informally, Claim (C1) states that the FDP of the WAMDF is asymptotically always larger than the FDP of its unadaptive counterpart and is referred to as the *asymptotically less conservative* claim. Claim (C2) states that the WAMDF has asymptotic FDP that is less than or equal to α and is referred to as the *asymptotic FDP control* claim. Claim (C3) is the α -*exhaustive* claim and states that the asymptotic FDP of the WAMDF is equal to α under a DU distribution. Theorem 5 provides sufficient conditions for each claim.

Theorem 5. Fix $\alpha \in (0, 1)$ and suppose that (A2) and (A4) - (A6) are satisfied. Then Claim (C1) holds. Claim (C2) holds if, additionally, $\mu_0 \leq 1$. Claim (C3) holds for $0 < \alpha \leq FDP_\infty(u)$ if, additionally, $\mu_0 = 1$.

Observe that asymptotic FDP control (C2) and α -exhaustion (C3) depend on the unobservable value of μ_0 , which will necessarily depend on the weighting scheme at hand. In particular, FDP control occurs when $\mu_0 \leq 1$ while α -exhaustion occurs when $\mu_0 = 1$. To motivate the next Theorem, which deals with μ_0 , recall in Example 1 that an optimal fixed- t weight is increasing in p_m . In other words, an optimal weight under Model 1 is positively correlated with θ_m , which indicates the state of H_m . It turns out that this positive correlation condition is enough to imply that $\mu_0 \leq 1$ and hence is useful for verifying

asymptotic FDP control. In fact, if weights are uncorrelated with the states of the null hypotheses, then $\mu_0 = 1$ and α -exhaustion is achieved. Examples of such weighting schemes are provided later.

Theorem 6. *Suppose that $(W_{m,M}, \theta_{m,M}), m \in \mathcal{M}$ are identically distributed random vectors each with support $\mathbb{R}^+ \times \{0, 1\}$ and with $E[W_{m,M}] = 1$ and $E[\theta_{m,M}] \in (0, 1)$. Define,*

$$\bar{W}_{0,M} = \frac{\sum_{m \in \mathcal{M}} (1 - \theta_{m,M}) W_{m,M}}{\sum_{m \in \mathcal{M}} (1 - \theta_{m,M})}$$

whenever $\theta_M \neq \mathbf{1}_M$ and take $\bar{W}_{0,M} = 1$ otherwise. Further assume that $\bar{W}_{0,M} \rightarrow \mu_0$ almost surely. Then $\mu_0 \leq 1$ if $\text{Cov}(W_{m,M}, \theta_{m,M}) \geq 0$ and $\mu_0 = 1$ if $\text{Cov}(W_{m,M}, \theta_{m,M}) = 0$.

Before focusing on the asymptotically optimal WAMDF, we first provide Corollary 1, which is now easily established using Theorems 5 and 6. It states that Claims (C1) - (C3) are satisfied for the unweighted adaptive (linear step-up) procedure defined in Storey, Taylor and Siegmund (2004), which is denoted $\delta(\hat{t}_{\alpha,M}^\lambda \mathbf{1}_M)$ here. That is $\delta(\hat{t}_{\alpha,M}^\lambda \mathbf{1}_M)$ is less conservative than $\delta(\hat{t}_{\alpha,M}^0 \mathbf{1}_M)$ and provides asymptotic α -exhaustive FDP control.

Corollary 1. *Suppose that (A4) - (A6) are satisfied and take $\mathbf{w}_M = \mathbf{1}_M$. Then for any fixed $\lambda \in (0, 1)$ and $0 < \alpha \leq a_0$, Claims (C1) - (C3) hold.*

This Corollary, in particular Claim (C3), suggests that the procedure in Storey, Taylor and Siegmund (2004) is competitive with α -exhaustive nonlinear procedures in Finner, Dickhaus and Roters (2009). The fact that a DU distribution is the least favorable among such (unweighted) adaptive linear step-up procedures under our weak dependence structure is also interesting as the search for least favorable distributions remains a challenging problem, especially when considering step-up procedures; step-up procedures fail for certain nonlinear procedures in Finner, Dickhaus and Roters (2009) because they utilize threshold $t = 1$ and consequently result in asymptotic FDP equal to a_0 . See, Finner, Dickhaus and Roters (2007); Roquain and Villers (2011); Finner, Gontscharuk and Dickhaus (2012) for more on nonlinear procedures and least favorable distributions, and see Tamhane, Liu and Dunnett (1998) for formal definitions of step-up, step-down and step-up-down procedures.

6.2. Asymptotically optimal weights

6.2.1. FDP control

We first verify that the conditions allowing for the WAMDF to provide less conservative asymptotic FDP control (Claims (C1) and (C2)) are satisfied under Model 1, even if the asymptotically optimal weights are perturbed or “noisy”. As in the previous subsection, weight vectors and elements of weight vectors are

indexed by M to facilitate asymptotic arguments. Further, we sometimes write $\bar{t}_M(k_M^*) = \bar{t}_M(k_M^*, \mathbf{p}, \gamma)$ for brevity.

Perturbed weights are simulated by multiplying each asymptotically optimal weight by a positive random variable U_m so that, by the law of iterated expectation, perturbed weights are positively correlated with asymptotically optimal weights. They are formally defined by

$$\tilde{w}_{m,M}(k_M^*, \mathbf{p}, \gamma) = U_m w_{m,M}(k_M^*, \mathbf{p}, \gamma) \quad (18)$$

for each m . For short, a perturbed weight is often denoted by $\tilde{w}_{m,M}$ and the vector of perturbed weights is denoted by $\tilde{\mathbf{w}}_M(k_M^*, \mathbf{p}, \gamma)$ or simply by $\tilde{\mathbf{w}}_M$. To allow for (A2) to be satisfied we assume that each triplet (U_m, γ_m, p_m) has joint distribution satisfying $0 \leq U_m t_m(k_M^*/p_m, \gamma_m) \leq 1$ almost surely. Further, assume that $E[U_m | \mathbf{p}, \gamma] = 1$ for each m so that perturbed weights have mean 1. It should be noted $\tilde{\mathbf{w}}_M = \mathbf{w}_M^*$ if $U_m = 1$ for each m (almost surely). Hence, results regarding perturbed weights immediately carry over to asymptotically optimal weights.

Theorem 7 is formally stated below. Note that we assume $\Pr(p_m \leq 1 - \alpha) = 1$ in Model 1 so that $\alpha \leq 1 - p_{(M)}$ for every M with probability 1, which implies that asymptotically optimal weights and their perturbed versions exist; see Theorem 2.

Theorem 7. *Suppose that $\Pr(p_m \leq 1 - \alpha) = 1$, take $\lambda_M = \bar{t}_M(k_M^*)$, and consider perturbed weights $\tilde{\mathbf{w}}_M$. Under Model 1 and (A1), (A2) and (A4) - (A6) are satisfied and $\mu_0 \leq 1$. Hence the conditions of Theorem 4 are satisfied and Claims (C1) and (C2) hold.*

6.2.2. Optimality

In this subsection, we formally describe what is meant by “asymptotically optimal” and also provide some examples of α -exhaustive weighting schemes. To motivate the first Theorem, recall that the asymptotically optimal weights in the Weight Selection Procedure are equivalent to optimal fixed- t weights with $t = \bar{t}_M(k_M^*)$. However, the asymptotically optimal WAMDF utilizes asymptotic threshold $t_{\alpha, \infty}^\lambda$ (see Theorem 4). The next theorem states that $\bar{t}_M(k_M^*) \rightarrow t_{\alpha, \infty}^\lambda$ almost surely, i.e. the asymptotically optimal weights are asymptotically equivalent to the optimal fixed- t weights corresponding to Theorem 1. Note that we choose $\lambda_M = \bar{t}_M(k_M^*)$ in the asymptotically optimal WAMDF. Theorem 8 need not apply for other specifications of λ_M .

Theorem 8. *Suppose that $\Pr(p_m \leq 1 - \alpha) = 1$ and take $\lambda_M = \bar{t}_M(k_M^*)$. Then under Model 1 and (A1), $\bar{t}_M(k_M^*) \rightarrow t_{\alpha, \infty}^\lambda$ almost surely.*

The other notion of optimality considered in this paper is the notion of α -exhaustion. The next two corollaries illustrate that Claims (C1) - (C3) (note in particular the α -exhaustion Claim (C3)) are satisfied in wide variety of settings. Corollary 2 states that a WAMDF is α -exhaustive in a worst case scenario

weighting scheme, i.e. when weights are generated independent of θ_M . This is perhaps not surprising in light of Corollary 1, where we saw that even choosing $w_M = \mathbf{1}_M$ results in α -exhaustion.

Corollary 2. *Under Model 1 and (A1) - (A2), if w_M are mutually independent weights and independent of θ_M with $E[w_{m,M}] = 1$, then Claims (C1) - (C2) hold for $\alpha \in (0, 1)$ and Claim (C3) holds for $0 < \alpha \leq FDP_\infty(u)$.*

Recall in Theorem 7 that optimal weights and their perturbed versions provide less conservative asymptotic FDP control (Claims (C1) and (C2)) under Model 1 and (A1). In Corollary 3, we see that if $p_i = p_j$ for every i, j , then optimal weights and their perturbed versions allow for α -exhaustion as well. This setting arises in practice whenever no information for distinguishing prior probabilities \mathbf{p} from one another is available, but some information for distinguishing effect sizes from one another may be available. For an illustration see Section 8. See also Spjøtvoll (1972); Storey (2007); Peña, Habiger and Wu (2011).

Corollary 3. *Suppose that the conditions of Theorem 7 are satisfied and consider perturbed weights \tilde{w}_M . If additionally $p_i = p_j$ for every i, j then Claim (C3) holds for $0 < \alpha \leq FDP_\infty(u)$.*

It should be noted that α -exhaustion need not be achieved when $p_i \neq p_j$ in Model 1 for the asymptotically optimal WAMDF. Hence, though more powerful than its competitors, such as the α -exhaustive unweighted MDF $\delta(t_{\alpha,M}^\lambda \mathbf{1}_M)$, there is room for additional improvement. A similar phenomenon was observed in Genovese, Roeder and Wasserman (2006) in the unadaptive setting, and it was suggested that one potential route for improvement is to incorporate an estimate of μ_0 into the procedure. However, it is not clear how this objective could be accomplished without sacrificing FDP control, especially when weights may be perturbed. We leave this as future work.

7. Simulation

This section compares weighted adaptive MDFs to other MDFs in terms of power and FDP control via simulation. In particular, for each of $K = 1000$ replications, we generate $Z_m \stackrel{i.i.d.}{\sim} N(\theta_m \gamma_m, 1)$ for $m = 1, 2, \dots, 1000$ and compute $\delta(\hat{t}_{\alpha,M}^\lambda w_M)$, $\delta(\hat{t}_{\alpha,M}^0 w_M)$, $\delta(\hat{t}_{\alpha,M}^\lambda \mathbf{1}_M)$, and $\delta(\hat{t}_{\alpha,M}^0 \mathbf{1}_M)$ as in Example 1, where $\alpha = 0.05$ and $\lambda_M = \hat{t}_M(k_M^*, \mathbf{p}, \gamma)$. Recall these MDFs are referred to as weighted adaptive (WA), weighted unadaptive (WU), unweighted adaptive (UA), and unweighted unadaptive (UU) procedures, respectively. The WU procedure is akin to the weighted BH procedure in Genovese, Roeder and Wasserman (2006) while the UA procedure is the adaptive BH procedure in Storey, Taylor and Siegmund (2004). The UU procedure is the original BH procedure. The average FDP and average correct discovery proportion (CDP) is computed over the K replications for each procedure, where the CDP is defined by $CDP = \sum_{m \in \mathcal{M}_1} \delta_m / \max\{M_1, 1\}$.

In each simulation experiment, $\gamma_m \stackrel{i.i.d.}{\sim} Un(1, a)$ for $a = 1, 3, 5$, where $Un(1, a)$ denotes a uniform distribution over $(1, a)$. Observe that when $a = 1$ the effect

sizes are identical while when $a = 3$ or $a = 5$ they vary. In Simulation 1, we take $p_m = 0.5$ for each m and weighted procedures utilize asymptotically optimal weights. The data generating mechanism here is similar to settings considered in Spjøtvoll (1972), Storey (2007), and Peña, Habiger and Wu (2011) in that no distinguishing prior probabilities for the states of the H_m s are available and results in the WAMDF being both optimally weighted and α -exhaustive; see Corollary 3. This setting also arises in the analysis of the data in Table 1 in the next section. In Simulation 2, weighted procedures use asymptotically optimal weights as before and the effect sizes vary as before, but the prior probabilities now vary via $p_m \stackrel{i.i.d.}{\sim} Un(0, 1)$. Thus, the conditions of Claim (C3) are no longer satisfied and the WA procedure is optimally weighted but not α -exhaustive (recall the last paragraph of Section 6). In Simulation 3, data are generated according to the same mechanism as in Simulation 2, but asymptotically optimal weights are perturbed via $U_m w_{m,M}(k_M^*, \mathbf{p}, \boldsymbol{\gamma})$, where $U_m \stackrel{i.i.d.}{\sim} Un(0, 2)$; see expression (18). Hence, the WA procedure is no longer optimally weighted nor is it α -exhaustive, but the employed weights are positively correlated with optimal weights. Simulation 4 considers a worst case scenario in which weights are generated $w_{m,M} \stackrel{i.i.d.}{\sim} Un(0, 2)$. Even in this worst case scenario α -exhaustion is achieved for the WA procedure by Corollary 2. Note that the UA procedure is α -exhaustive in all four simulations due to Corollary 1 while the unadaptive procedures are never α -exhaustive. Results are summarized in Table 3.

The main important point is that the WA procedure dominates all other procedures as long as the employed weights are at least positively correlated with the optimal weights, and it performs nearly as well as other procedures otherwise. In particular, its FDP is less than or equal to 0.05 in all simulations and its average CDP is as large or larger than the CDP of all other procedures in the first three simulations. The WA procedure does have a smaller average CDP than the UA procedure in the worst case scenario (Simulation 4), as we might expect.

Now let us focus on Simulation 1 in more detail. First, observe that the FDP is increasing in a for both adaptive procedures. For example the FDP of each procedure is 0.021 when $a = 1$ but is 0.039 when $a = 5$. This is to be expected as both adaptive procedures are α -exhaustive (see Corollaries 1 and 3) and hence we should expect the FDP to be near 0.05 in high power settings, i.e. for large a . Additionally, the largest gain in power (in terms of the average CDP) of the weighted adaptive procedure over the unweighted adaptive procedure occurs when effect sizes are most heterogeneous. In particular, when $a = 5$ the average CDP of the WA procedure is 0.793 while the average CDP of the UA procedure is 0.761. When data are homogeneous ($a = 1$), the CDPs of the procedures are identical.

In Simulation 2, data generating mechanisms are even more heterogeneous as now the p_m s also vary. General conclusions regarding the CDP are the same, with the advantages of the weighted procedures over their unweighted counterparts being more pronounced. For example, the average CDP of the WAMDF for $\gamma_m \stackrel{i.i.d.}{\sim} Un(1, 5)$ increased from 0.793 to 0.814 when allowing p_m s to vary,

TABLE 3
The average CDP (FDP) for the UU, UA, WU, and WA procedures in Simulations 1 - 4.

Simulation 1			
	a		
	1	3	5
UU	0.007(0.021)	0.390(0.025)	0.709(0.025)
WU	0.007(0.021)	0.397(0.025)	0.731(0.025)
UA	0.007(0.021)	0.437(0.031)	0.761(0.039)
WA	0.007(0.021)	0.442(0.032)	0.793(0.039)

Simulation 2			
	a		
	1	3	5
UU	0.007(0.024)	0.390(0.025)	0.709(0.025)
WU	0.011(0.008)	0.434(0.014)	0.756(0.016)
UA	0.007(0.024)	0.430(0.031)	0.761(0.041)
WA	0.011(0.008)	0.473(0.018)	0.814(0.026)

Simulation 3			
	a		
	1	3	5
UU	0.007(0.023)	0.391(0.025)	0.709(0.025)
WU	0.013(0.007)	0.404(0.015)	0.719(0.016)
UA	0.007(0.023)	0.430(0.031)	0.757(0.039)
WA	0.013(0.007)	0.439(0.019)	0.774(0.027)

Simulation 4			
	a		
	1	3	5
UU	0.007(0.025)	0.391(0.025)	0.710(0.025)
WU	0.006(0.023)	0.354(0.025)	0.682(0.025)
UA	0.007(0.025)	0.425(0.030)	0.756(0.039)
WA	0.006(0.023)	0.387(0.030)	0.727(0.039)

while for the UA procedure the CDP is still 0.761. We also observe that for $a = 5$ the average FDP of the WA procedure is only 0.026 while the average FDP of the UA procedure is closer to 0.05; it is 0.039. This is to be expected because, even though the WAMDF will dominate the UA procedure in terms of the average CDP, the UA procedure is α -exhaustive while the WAMDF need not be in this setting.

Now consider non-optimal weights in Simulations 3 and 4. [Roeder and Wasserman \(2009\)](#) concluded that, in the unadaptive setting (the UU and WU procedures), weighted MDFs are robust with respect to weight misspecification in that they generally yield about as many or more rejected null hypotheses as unweighted procedures as long as weights are “reasonably guessed” and yield slightly less rejected null hypotheses when weights are poorly guessed. Simulations 3 and 4 confirm their results and further illustrates that the robustness property applies to adaptive procedures. For example, comparing the unadaptive procedures in Simulation 3, we see that the average CDP of the WU(UU) procedures are 0.013(0.007), 0.404(0.391), and 0.719(0.709) for $a = 1, 3, 5$, respectively. The average CDP of the WA(UA) procedure is 0.013(0.007), 0.439(0.430), and 0.774(0.757), for $a = 1, 3, 5$, respectively. That is, when weights are positively correlated with optimal weights, weighted procedures still perform slightly better than their unweighted counterparts. In the worst case scenario setting

in Simulation 4, where weights are independently generated, the FDP is still controlled by the WA procedure, but some loss in power over its unweighted counterpart is observed. For example, the CDP of the WA(UA) procedure is 0.006(0.007), 0.386(0.425), and 0.727(0.756) when $\gamma = 1, 3$, and 5, respectively, while the average FDP of the WA(UA) procedure is 0.025(0.023), 0.030(0.030), and 0.039(0.039) when $\gamma = 1, 3$, and 5, respectively. Again the α -exhaustion phenomenon for both adaptive procedures is observed in Simulation 4, as Corollaries 1 and 2 suggest.

8. Data analysis

Next we illustrate the method on the data set outlined in Table 1. Recall that the goal is to determine which bacteria are positively associated with shoot biomass via the testing of $H_m : \beta_{1m} = 0$ vs. $K_m : \beta_{1m} > 0$ for each m , where β_{1m} is the regression coefficient for regressing $\mathbf{Y}_m = (Y_{im}, i = 1, 2, \dots, 5)$ on $\mathbf{x} = (x_i, i = 1, 2, \dots, 5)$ with a log-linear model defined by $\log(\mu_{im}) = \beta_{0m} + \beta_{1m}x_i$ for μ_{im} the mean of Y_{im} . To define the problem within the context of Model 1, assume $\beta_{1m} = \beta\theta_m$ for some $\beta > 0$ so that the null hypothesis can be written $H_m : \theta_m = 0$. Further assume $\mathbf{Y}_m | (\sum_{i=1}^5 Y_{im} = n_m)$ has a multinomial distribution with sample size parameter n_m and probability vector $\mathbf{p}_m = (p_{im}, i = 1, 2, \dots, 5)$. Note that under the log linear model with $\beta_{1m} = \beta\theta_m$, $p_{im} = e^{\beta\theta_m x_i} / [\sum_{i=1}^5 e^{\beta\theta_m x_i}]$. Further, $p_{im} = 1/5$ if $\theta_m = 0$.

The p -value based on test statistic $T_m = \mathbf{x}^T \mathbf{Y}_m$ is defined $P_m = \Pr(T_m \geq t_m | \sum_{i=1}^5 Y_{im} = n_m, \theta_m = 0)$. See McCullagh and Nelder (1989) for details. The power function for $\delta_m(P_m; t_m) = I(P_m \leq t_m)$ is approximated with the power function expression in Example 1, where $\gamma_m = \sqrt{n_m}K(\beta)$ for $K(\beta)$ some positive constant depending on β . For details on this approximation see the Supplement. The important point is that effect sizes are proportional to $\sqrt{n_m}$. It is anticipated in Anderson and Habiger (2012) that roughly 1/2 of bacteria will be positively associated with shoot biomass and that, of these, about 1/2 will be detected. Thus, we choose $p_m = 1/2$ for each m in Model 1 and choose $K(\beta)$ so that average power $M^{-1} \sum_m \pi_{\gamma_m}(t_m(k_M^*/p_m, \gamma_m)) = 1/2$.

The asymptotically optimal weights and the (posited) power of the unweighted and weighted procedures, given by $\pi_{\gamma_m}(t)$ and $\pi_{\gamma_m}(tw_m^*)$ for $t = \bar{t}_M(k_M^*)$, respectively, are in Figure 2. Observe that $w_m^* > 1$ for the first six points (which actually represents 558 of the 778 hypothesis tests with the smallest sample sizes - many of the n_m 's, and consequently the w_m^* s, are identical and hence each point represents multiple weights) and resulted in increased power; see the right panel of Figure 2. Optimal weights for the remaining 220 hypotheses are less than 1 and can even be near 0. However, in these settings, the power of the weighted decision function is still very near 1. In short, the optimal strategy borrows weight from tests where the power is anticipated to be near 1, even if the weight is near 0, and redistributes it to tests where increasing the weight can have a more substantial effect on power. Indeed, for $\alpha = 0.05$ the asymptotically optimal WAMDF resulted in 38 discoveries whereas the unweighted

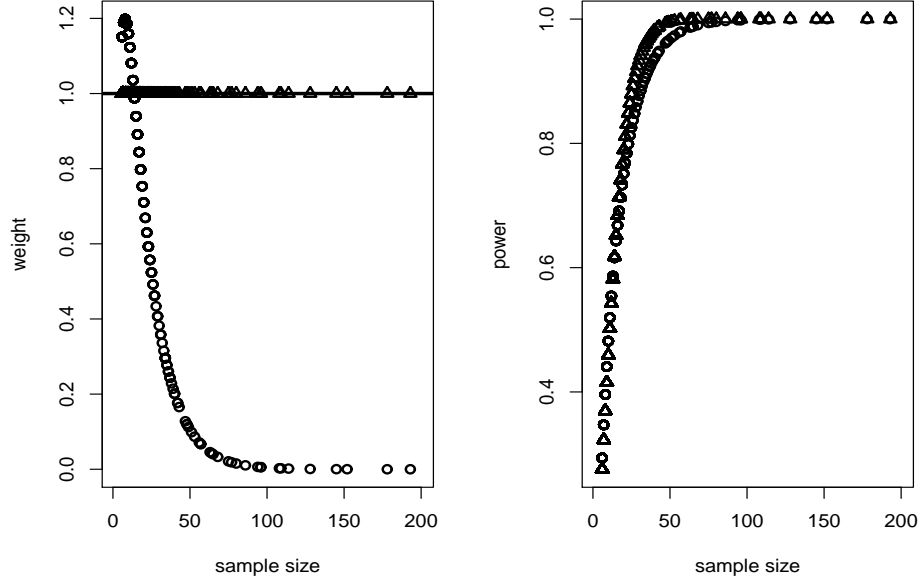


FIG 2. The left panel plots w_m^* and $w_m = 1$ vs. γ_m , denoted by o and \triangle , respectively. The right panel plots $\pi_{\gamma_m}(tw_m^*)$ and $\pi_{\gamma_m}(t)$ vs. γ_m , denoted by o and \triangle , respectively, where $t = \bar{t}_M(k_M^*)$ is computed as in the Weight Selection Procedure. Not depicted, for readability, are weights and posited power for four H_m s with sample sizes 333, 457, 883, 911. These posited powers and weights were all 1 and 0, respectively.

adaptive procedure yielded only 32 discoveries. While the FDP is unobservable, Corollary 3 suggests that it is less than or equal to 0.05, even if the assumptions under which optimal weights were computed need not be correct.

9. Concluding Remarks

Efforts to improve upon the original BH procedure have focused on 1) controlling the FDR at a level nearer α or 2) exploiting potential heterogeneity across tests. This paper combined these ideas using a weighted decision theoretic framework and showed that the resulting procedure is more powerful than procedures which only consider 1) or 2), but not both. Specifically, we have provided weighted adaptive multiple decision functions that satisfy the α -exhaustive optimality criterion considered in Finner, Dickhaus and Roters (2009), but allow for further improvements via an optimal weighting scheme that incorporates heterogeneity. For example, in Corollary 1 we saw that the unweighted adaptive procedure is α -exhaustive. However, its weighted version is even more powerful when data are heterogeneous and optimal weights are used; see Theorems 1 and 8 and the CDP

of the weighted adaptive (WA) vs. the unweighted adaptive (UA) procedures in Simulations 1 and 2 in Section 7. In fact, even when weights are not optimal, we saw in Simulation 3 that some gain in power is achieved.

Finite sample results and asymptotic results in this paper are valid under independence and weak dependence conditions, respectively. Benjamini and Yekutieli (2001) showed that the unweighted unadaptive BH procedure provides (finite) FDR control under a certain positive dependence structure, and can be modified to control the FDR for arbitrary dependence. It would be interesting to study the performance of weighted adaptive procedures under arbitrary dependence. However, obtaining finite sample analytical results, even in the unweighted adaptive setting, appears to be very challenging, especially under arbitrary dependence. See Guo and Sarkar (2012) for some results regarding unweighted procedures under a block dependence-type structure. As for large sample results, Fan, Han and Gu (2012); Desai and Storey (2012) provide techniques for transforming test statistics so that they are weakly dependent. It may be possible to construct weighted adaptive MDFs which make use of these transformed statistics, but this also requires further study.

Optimal weighting schemes were derived assuming γ and \mathbf{p} were known, i.e. we assumed that the nature and degree of heterogeneity is known. In some settings, consistent estimates of these parameters are readily available and WA procedures are easily implemented. See Cai and Sun (2009), Hu, Zhao and Zhou (2010), or Sun and McLain (2012) for examples. In other settings, consistent estimates may not be available and the nature and degree of heterogeneity may not be precisely known. However, asymptotic FDP control is still provided even if heterogeneity is poorly modeled/posited (see, in particular, Theorem 7 and Corollary 2) and as long as the weighting scheme is reasonable, then some gain in power should be anticipated (see Simulation 3). Hence, heterogeneity should be incorporated into the multiple testing procedure even if only a reasonable guess for unknown parameters may be available, as in Section 8, where effect sizes were posited based on the assumption that tests with larger sample sizes are more powerful.

Other estimators for M_0 could be considered. For example, it is possible to use the unweighted estimator from Storey, Taylor and Siegmund (2004) in the WAMDF. One reason for using the weighted estimator is that this simplifies analytical arguments and the implementation of the WAMDF. In particular, as we saw in Section 4, it allows for the WAMDF to be implemented by applying the adaptive BH procedure in Storey, Taylor and Siegmund (2004) to the weighted p -values, and facilitates martingale arguments for finite sample results. Further, under a DU distribution and the dependence structure in (A3), it can be verified that the weighted estimator performs better than the unweighted estimator in that both estimators are consistent but $Var(\hat{M}_0(\lambda \mathbf{1})) \geq Var(\hat{M}_0(\lambda \mathbf{w}))$. The latter claim is due to Hoeffding (1956), where it was shown that the variance of the sum of n independent Bernoulli random variables with average success probability $n^{-1}(p_1 + p_2 + \dots + p_n) = p$ is maximized when $p_1 = p_2 = \dots = p_n = p$. Of course, this does not imply that the weighted estimator will always perform better than the unweighted estimator. However, a more detailed assessment of

$\hat{M}_0(\lambda \mathbf{w})$, though warranted, is beyond the scope of the current manuscript.

The robustness properties of the asymptotically optimal WAMDF suggest that simpler weighting schemes may still allow for improvements over unweighted adaptive procedures and hence, in some settings, may be an attractive alternative to the more complex asymptotically optimal weighting scheme defined in Section 3. For example, [Genovese, Roeder and Wasserman \(2006\)](#) considered a binary weighting scheme where each hypothesis was given one of two possible weights and demonstrated that the resulting weighted (unadaptive) procedure could still be more powerful than its unweighted counterpart. It would be interesting to study simpler weighting schemes within the context of adaptive procedures.

References

- ANDERSON, M. and HABIGER, J. (2012). Characterization and Identification of Productivity-Associated Rhizobacteria in Wheat. *Appl. Environ. Microbiol.* **78** 4434 - 444.
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. [MR1325392 \(96d:62143\)](#)
- BENJAMINI, Y. and HOCHBERG, Y. (2000). On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. *J. Educ. Behav. Stat.* **25** 60–83.
- BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. [MR2261438 \(2007j:62024\)](#)
- BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. [MR1869245 \(2002i:62135\)](#)
- BILLINGSLEY, P. (1999). *Convergence of probability measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. [MR1700749 \(2000e:60008\)](#)
- BLANACHAR, G. and ROQUAIN, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.* **2** 963–992. [MR2448601 \(2009k:62184\)](#)
- BLANCHARD, G. and ROQUAIN, E. (2009). Adaptive FDR control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2831.
- CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481. [MR2597000 \(2011d:62020\)](#)
- CAO, H., SUN, W. and KOSOROK, M. R. (2013). The optimal power puzzle: scrutiny of the monotone likelihood ratio assumption in multiple testing. *Biometrika* **100** 495–502. [MR3068449](#)
- DESAI, K. H. and STOREY, J. D. (2012). Cross-dimensional inference of dependent high-dimensional data. *J. Amer. Statist. Assoc.* **107** 135–151. [MR2949347](#)
- DOOB, J. (1953). *Stochastic Processes*. John Wiley and Sons, New York.
- DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple testing procedures with applications to genomics*. *Springer Series in Statistics*. Springer, New York. [MR2373771 \(2009j:62004\)](#)
- EFRON, B. (2008). Microarrays, Empirical Bayes and the Two-Group Model. *Statistical Science* **23** 1–22.
- EFRON, B. (2010). *Large-scale inference*. *Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge University Press, Cambridge. Empirical Bayes methods for estimation, testing, and prediction. [MR2724758 \(2012a:62006\)](#)
- EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–

1160. [MR1946571](#)
- FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. [MR3010887](#)
- FINNER, H., DICKHAUS, T. and ROTERS, M. (2007). Dependency and false discovery rate: asymptotics. *Ann. Statist.* **35** 1432–1455. [MR2351092 \(2008j:62020\)](#)
- FINNER, H., DICKHAUS, T. and ROTERS, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37** 596–618. [MR2502644 \(2010f:62198\)](#)
- FINNER, H., GONTSCHARUK, V. and DICKHAUS, T. (2012). False discovery rate control of step-up-down tests with special emphasis on the asymptotically optimal rejection curve. *Scand. J. Stat.* **39** 382–397. [MR2927031](#)
- GAVRILOV, Y., BENJAMINI, Y. and SARKAR, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.* **37** 619–629. [MR2502645 \(2010f:62199\)](#)
- GENOVESE, C., ROEDER, K. and WASSERMAN, L. (2006). False discovery control with p -value weighting. *Biometrika* **93** 509–524. [MR2261439 \(2008c:62018\)](#)
- GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristic and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 499–517.
- GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. [MR2065197 \(2005k:62149\)](#)
- GUO, W. and SARKAR, S. (2012). Adaptive controls of FWER and FDR under block dependence. *available at web.njit.edu/wguo*.
- HABIGER, J. (2012). A method for modifying multiple testing procedures. *J. Statist. Plann. Inference* **142** 2227–2231. [MR2903425](#)
- HABIGER, J. and PEÑA, E. (2011). Randomized p -values and nonparametric procedures in multiple testing. *J. Nonparametr. Stat.* **23** 583–604.
- HABIGER, J. D. and PEÑA, E. A. (2014). Compound p -value statistics for multiple testing procedures. *J. Multivariate Anal.* **126** 153–166. [MR3173088](#)
- HOEFFDING, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721. [MR0080391 \(18,240f\)](#)
- HU, J. X., ZHAO, H. and ZHOU, H. H. (2010). False discovery rate control with groups. *J. Amer. Statist. Assoc.* **105** 1215–1227. [MR2752616 \(2012b:62252\)](#)
- LIANG, K. and NETTLETON, D. (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74** 163–182.
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized linear models. Monographs on Statistics and Applied Probability*. Chapman & Hall, London Second edition [of [MR0727836](#)]. [MR3223057](#)
- PEÑA, E. A., HABIGER, J. D. and WU, W. (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist.* **39** 556–583. [MR2797856 \(2012b:62055\)](#)

- PEÑA, E., HABIGER, J. and WU, W. (2014). Classes of multiple decision functions strongly controlling FWER and FDR. *Metrika* 1-33.
- ROEDER, K. and WASSERMAN, L. (2009). Genome-wide significance levels and weighted hypothesis testing. *Statist. Sci.* **24** 398–413. [MR2779334 \(2011k:62307\)](#)
- ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3** 678–711. [MR2521216 \(2010k:62307\)](#)
- ROQUAIN, E. and VILLERS, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612. [MR2797857 \(2012b:62255\)](#)
- SARKAR, S. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* **35** 2405–2420. [MR2382652 \(2009e:62260\)](#)
- SCHWEDER, T. and SPJOTVOLL, E. (1982). Plots of P-Values to Evaluate Many Tests Simultaneously. *Biometrika* **69** pp. 493-502.
- SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. [MR897872 \(88d:62043\)](#)
- SPJOTVOLL, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Statist.* **43** 398–411. [MR0301871 \(46 #1026\)](#)
- STEVENS, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika* **37** 117–129. [MR0035955 \(12,37e\)](#)
- STOREY, J. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat Methodol.* **64** 479 - 498.
- STOREY, J. (2003). The positive false discovery rate: a Bayesian interpretation and the q-Value. *Ann. Statist.* **31** 2012 - 2035.
- STOREY, J. D. (2007). The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 347–368. [MR2323757](#)
- STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. [MR2035766 \(2004k:62056\)](#)
- SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. [MR2411657](#)
- SUN, W. and McLAIN, A. C. (2012). Multiple testing of composite null hypotheses in heteroscedastic models. *J. Amer. Statist. Assoc.* **107** 673–687. [MR2980076](#)
- TAMHANE, A. C., LIU, W. and DUNNETT, C. W. (1998). A generalized step-up-down multiple test procedure. *Canad. J. Statist.* **26** 353–363. [MR1648451](#)

10. Supplementary Material

10.1. Proofs of results in Section 3

PROOF OF THEOREM 1. Setting up the Lagrangian

$$L(\mathbf{t}, k) = \pi(\mathbf{t}, \mathbf{p}, \gamma) - k \left[\left(\sum_{m \in \mathcal{M}} t_m \right) - Mt \right]$$

and taking derivative with respect to t_m and setting it equal to 0 yields equation (7). Now, recall we denote the solution to equation (7) with respect to t_m by $t_m(k/p_m, \gamma_m)$ and observe $k \mapsto t_m(k/p_m, \gamma_m)$ is continuous and strictly decreasing in k with $\lim_{k \rightarrow \infty} t_m(k/p_m, \gamma_m) = 0$ and $\lim_{k \downarrow 0} t_m(k/p_m, \gamma_m) = 1$ by (A1). Thus, $\bar{t}_M(k, \mathbf{p}, \gamma) = M^{-1} \sum_{m \in \mathcal{M}} t_m(k/p_m, \gamma_m)$ is continuous and strictly decreasing in k with $\lim_{k \rightarrow \infty} \bar{t}_M(k, \mathbf{p}, \gamma) = 0$ and $\lim_{k \downarrow 0} \bar{t}_M(k, \mathbf{p}, \gamma) = 1$. Hence, there exists a unique k satisfying $\bar{t}_M(k, \mathbf{p}, \gamma) = t$ for any $t \in (0, 1)$ and hence a unique collection $[t_m(k/p_m, \gamma_m), m \in \mathcal{M}]$.

To show that the solution is a maximum, it suffices to show that the sequence of the determinants of the principal minors of the bordered hessian matrix, evaluated at the solution, alternates in sign. The j th principle minor of the bordered Hessian matrix is

$$\mathbf{H}_j = \begin{bmatrix} 0 & \mathbf{1}_j \\ \mathbf{1}_j^T & \mathbf{D}_j \end{bmatrix}$$

where \mathbf{D}_j is a $j \times j$ diagonal matrix with diagonal elements $d_m = \pi''_{\gamma_m}(t_m)$ and $\mathbf{1}_j$ is a vector of 1s of length j . Note that $d_m < 0$ at the solution due to (A1). Now, observe that $|\mathbf{H}_1| = -1 < 0$ where $|\cdot|$ denotes the determinant, and for $j \geq 2$, we have the recursive relation

$$|\mathbf{H}_j| = d_j |\mathbf{H}_{j-1}| + (-1)^j \prod_{m=1}^{j-1} (-d_m). \quad (\text{S1})$$

Because $d_j < 0$, for j an even (odd) integer each expression on the righthand side of equation (S1) is positive (negative). Hence $\{|\mathbf{H}_j|, j = 1, 2, \dots\}$ alternates in sign. The above arguments hold with probability 1. \square

PROOF OF THEOREM 2. Observe that $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma))$ is continuous in k under (A1). Hence, it suffices to show that $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma))$ takes on values 0 and $1 - p_{(M)}$ by the Mean Value Theorem. We first show that

$$\lim_{k \downarrow 0} \widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) \geq 1 - p_{(M)}.$$

Observe that (A1) implies $t_m \leq \pi_{\gamma_m}(t_m) \leq 1$ for $t_m \in [0, 1]$ and hence

$$\bar{t}_M(k, \mathbf{p}, \gamma) \leq \bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) \leq M^{-1} \left[\sum_{m \in \mathcal{M}} (1 - p_m) t_m(k/p_m, \gamma_m) + p_m \right]. \quad (\text{S2})$$

The inequalities in (S2) imply

$$\begin{aligned} \widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) &= \frac{\sum_{m \in \mathcal{M}} [1 - G_m(t_m(k/p_m, \gamma_m))]}{\sum_{m \in \mathcal{M}} [1 - t_m(k/p_m, \gamma_m)]} \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} \\ &\geq \frac{\sum_{m \in \mathcal{M}} [1 - p_m] [1 - t_m(k/p_m, \gamma_m)]}{\sum_{m \in \mathcal{M}} [1 - t_m(k/p_m, \gamma_m)]} \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} \\ &\geq (1 - p_{(M)}) \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))}, \end{aligned}$$

which converges to $1 - p_{(M)}$ as $k \downarrow 0$ if

$$\frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} \rightarrow 1 \quad (\text{S3})$$

as $k \downarrow 0$. To verify (S3), observe that $\lim_{k \downarrow 0} t_m(k/p_m, \gamma_m) = 1$ by (A1) and hence $\bar{t}_M(k, \mathbf{p}, \gamma) \rightarrow 1$ as $k \downarrow 0$. This, along with the inequalities in (S2), imply $\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) \rightarrow 1$ as $k \downarrow 0$ and hence (S3) is satisfied.

Now if

$$\lim_{k \rightarrow \infty} \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} = 0, \quad (\text{S4})$$

then by the first inequality in (S2) and the definition of $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma))$

$$\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma)) \leq \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} \rightarrow 0$$

as $k \rightarrow \infty$ and the proof would be complete. Hence, it suffices to show (S4). But because $t_m(k/p_m, \gamma_m) \downarrow 0$ as $k \rightarrow \infty$ and $\pi'_{\gamma_m}(t_m) \rightarrow \infty$ as $t_m \downarrow 0$ by (A1), we have

$$\frac{\pi_{\gamma_m}(t_m(k/p_m, \gamma_m))}{t_m(k/p_m, \gamma_m)} \rightarrow \infty$$

as $k \rightarrow \infty$ by Hôpital's rule. Further for $a_m, b_m, m \in \mathcal{M}$ any positive constants,

$$\frac{\sum_{m \in \mathcal{M}} a_m}{\sum_{m \in \mathcal{M}} b_m} = \sum_{m \in \mathcal{M}} \frac{a_m}{b_m} \left(\frac{b_m}{\sum_{m \in \mathcal{M}} b_m} \right) \geq \min \left\{ \frac{a_m}{b_m}, m \in \mathcal{M} \right\}.$$

Hence,

$$A(k) \equiv \frac{\sum_{m \in \mathcal{M}} \pi_{\gamma_m}(t_m(k/p_m, \gamma_m))}{\sum_{m \in \mathcal{M}} t_m(k/p_m, \gamma_m)} \geq \min \left\{ \frac{\pi_{\gamma_m}(t_m(k/p_m, \gamma_m))}{t_m(k/p_m, \gamma_m)}, m \in \mathcal{M} \right\} \rightarrow \infty$$

as $k \rightarrow \infty$ which implies

$$\begin{aligned} \frac{\bar{t}_M(k, \mathbf{p}, \gamma)}{\bar{G}_M(\mathbf{t}(k, \mathbf{p}, \gamma))} &= \left[\sum_{m \in \mathcal{M}} \frac{(1 - p_m) t_m(k/p_m, \gamma_m)}{\bar{t}_M(k, \mathbf{p}, \gamma)} + \frac{p_m \pi_{\gamma_m}(t_m(k/p_m, \gamma_m))}{\bar{t}_M(k, \mathbf{p}, \gamma)} \right]^{-1} \\ &\leq [M(1 - p_{(M)}) + Mp_{(1)}A(k)]^{-1} \rightarrow 0 \end{aligned}$$

as $k \rightarrow \infty$, where $p_{(1)} \equiv \min\{\mathbf{p}\}$. Note that the above arguments hold with probability 1. \square

10.2. Proofs of results in Section 5

PROOF OF LEMMA 1. The proof uses techniques from the proofs of Theorem 3 in Storey, Taylor and Siegmund (2004) and Theorem 9 in Peña, Habiger and Wu (2011, 2014). First, observe that because $u = \lambda$, $0 \leq \hat{t}_\alpha^\lambda \leq \lambda$ by definition and that if $\hat{t}_\alpha^\lambda = 0$ then $FDR(\hat{t}_\alpha^\lambda \mathbf{w}) = 0$ trivially. Let us focus on the setting where $0 < \hat{t}_\alpha^\lambda \leq \lambda$. By the definition of \hat{t}_α^λ , $\widehat{FDP}^\lambda(\hat{t}_\alpha^\lambda \mathbf{w}) \leq \alpha$ which gives $R(\hat{t}_\alpha^\lambda) \geq \hat{M}_0(\lambda \mathbf{w}) \hat{t}_\alpha^\lambda / \alpha$ by the definition of $\widehat{FDP}^\lambda(\cdot)$. Hence,

$$FDR(\hat{t}_\alpha^\lambda \mathbf{w}) = E \left[\frac{V(\hat{t}_\alpha^\lambda \mathbf{w})}{R(\hat{t}_\alpha^\lambda \mathbf{w})} \right] \leq E \left[\alpha \frac{1}{\hat{M}_0(\lambda \mathbf{w})} \frac{V(\hat{t}_\alpha^\lambda \mathbf{w})}{\hat{t}_\alpha^\lambda} \right] \quad (\text{S5})$$

$$\leq E \left[\frac{\alpha}{\hat{M}_0(\lambda \mathbf{w})} \frac{V(\lambda \mathbf{w})}{\lambda} \right], \quad (\text{S6})$$

where (S6) is established as follows. First, if $\hat{t}_\alpha^\lambda = \lambda$, it is true trivially. Now suppose that $0 < \hat{t}_\alpha^\lambda < \lambda$. Define filtration $\mathcal{F}_t = \sigma\{\boldsymbol{\delta}(s\mathbf{w}), 0 < t \leq s \leq \lambda\}$ and observe that \hat{t}_α^λ is a stopping time with respect to \mathcal{F}_t (with time running backwards). Further, for $0 < t \leq \lambda$, $V(t\mathbf{w})/t$ is a reverse martingale with respect

to \mathcal{F}_t . This can be verified by noting that for $0 < s \leq t \leq \lambda$

$$\begin{aligned}
E \left[\frac{V(s\mathbf{w})}{s} \middle| \mathcal{F}_t \right] &= \frac{1}{s} \sum_{m \in \mathcal{M}_0} E[\delta_m(sw_m) | \mathcal{F}_t] \\
&= \frac{1}{s} \sum_{m \in \mathcal{M}_0} \delta_m(tw_m) E[\delta_m(sw_m) | \delta_m(tw_m) = 1, \mathcal{F}_t] \\
&= \frac{1}{s} \sum_{m \in \mathcal{M}_0} \delta_m(tw_m) E[\delta_m(sw_m) | \delta_m(tw_m) = 1] \\
&= \frac{1}{s} \sum_{m \in \mathcal{M}_0} \delta_m(tw_m) \frac{sw_m}{tw_m} \\
&= \sum_{m \in \mathcal{M}_0} \frac{\delta_m(tw_m)}{t} \\
&= \frac{V(t\mathbf{w})}{t},
\end{aligned}$$

where first equality follows by the definition of $V(\cdot)$ and the second is due to the fact that $\delta_m(sw_m) = 0$ if $\delta_m(tw_m) = 0$ by the NS assumptions. The third equality is satisfied due to (A3). The forth equality follows by the fact that $\Pr([\delta_m(sw_m) = 1] \cap [\delta_m(tw_m) = 1]) = E[\delta_m(sw_m)] = sw_m$ for $m \in \mathcal{M}_0$ and $s \leq \lambda$ under the NS assumptions and under (A2). The forth and fifth equalities follow from some algebra and the definition of $V(\cdot)$, respectively. Hence, by the law of iterated expectation and the Optional Stopping Theorem (Doob, 1953)

$$\begin{aligned}
E \left[\frac{\alpha}{\hat{M}_0(\lambda\mathbf{w})} \frac{V(\hat{t}_\alpha^\lambda \mathbf{w})}{\hat{t}_\alpha^\lambda} \right] &= E \left\{ \frac{\alpha}{\hat{M}_0(\lambda\mathbf{w})} E \left[\frac{V(\hat{t}_\alpha^\lambda \mathbf{w})}{\hat{t}_\alpha^\lambda} \middle| \mathcal{F}_\lambda \right] \right\} \\
&= E \left[\frac{\alpha}{\hat{M}_0(\lambda\mathbf{w})} \frac{V(\lambda\mathbf{w})}{\lambda} \right].
\end{aligned}$$

Hence, we have established (S6).

Now, note that $M - R(\lambda\mathbf{w}) = M_0 - V(\lambda\mathbf{w}) + [M_1 - \sum_{\mathcal{M}_1} \delta_m(\lambda w_m)] \geq M_0 - V(\lambda\mathbf{w})$. Further observe that $V \mapsto V(\lambda\mathbf{w})/[M_0 - V(\lambda\mathbf{w}) + 1]$ is convex. Hence, by Theorem 3 in Hoeffding (1956) and with $p = \lambda\bar{w}_0$

$$\begin{aligned}
E \left[\frac{V(\lambda\mathbf{w})}{M_0 - V(\lambda\mathbf{w}) + 1} \right] &\leq \sum_{k=0}^{M_0} \frac{k}{M_0 - k + 1} \binom{M_0}{k} p^k (1-p)^{M_0-k} \\
&= \frac{p}{1-p} (1 - p^{M_0}).
\end{aligned}$$

The last equality follows from basic calculations. Thus,

$$\begin{aligned} E \left[\alpha \frac{1}{\hat{M}_0(\lambda \mathbf{w})} \frac{V(\lambda \mathbf{w})}{\lambda} \right] &= E \left[\alpha \frac{(1-\lambda)}{M - R(\lambda \mathbf{w}) + 1} \frac{V(\lambda \mathbf{w})}{\lambda} \right] \\ &\leq \alpha \frac{(1-\lambda)}{\lambda} E \left[\frac{V(\lambda \mathbf{w})}{M_0 - V(\lambda \mathbf{w}) + 1} \right] \\ &= \alpha \frac{(1-\lambda)}{\lambda} \frac{p}{1-p} (1 - p^{M_0}). \end{aligned}$$

The result follows by plugging $\lambda \bar{w}_0$ in for p in the last expression. \square

PROOF OF THEOREM 3. From Lemma 1 and because $\bar{w}_0 \leq w_{(M)}$,

$$FDR(\hat{t}_{\alpha^*}^\lambda \mathbf{w}) \leq \alpha^* \bar{w}_0 \frac{1-\lambda}{1-\lambda \bar{w}_0} = \alpha \frac{\bar{w}_0}{w_{(M)}} \frac{1-\lambda w_{(M)}}{1-\lambda \bar{w}_0} \leq \alpha. \quad \square$$

10.3. Proofs of results in Section 6

Before proving Theorem 4 the following Glivenko-Cantelli-type Lemma regarding the uniform convergence of the FDP estimators and the FDP is presented. For similar results in the unweighted adaptive setting see Theorem 6 in Storey, Taylor and Siegmund (2004) or see the proof of Theorem 2 in Genovese, Roeder and Wasserman (2006) for the weighted, but unadaptive, setting. See also Finner, Dickhaus and Roters (2009); Fan, Han and Gu (2012) and references therein for additional results on almost sure convergence of the FDP.

Lemma 2. Fix $\delta \in (0, u)$. Under (A2) and (A4) - (A6),

$$\sup_{\delta \leq t \leq u} |\widehat{FDP}_M^0(t \mathbf{w}_M) - FDP_\infty^0(t)| \rightarrow 0,$$

$$\sup_{\delta \leq t \leq u} |\widehat{FDP}_M^\lambda(t \mathbf{w}_M) - FDP_\infty^\lambda(t)| \rightarrow 0,$$

and

$$\sup_{\delta \leq t \leq u} |FDP_M(t \mathbf{w}_M) - FDP_\infty(t)| \rightarrow 0$$

almost surely.

Proof. In what follows we denote $\max\{R(t \mathbf{w}_M), 1\}$ by $R(t \mathbf{w}_M)$ for short. Observe $R(t \mathbf{w}_M)$ is nondecreasing in t almost surely by the NS assumptions and

$G(t)$ is strictly increasing in t for $0 \leq t \leq u$ by (A6). Hence, for any $\delta \in (0, u)$,

$$\begin{aligned} \sup_{\delta \leq t \leq u} \left| \widehat{FDP}_M^0(t\mathbf{w}_M) - FDP_\infty^0(t) \right| &= \sup_{\delta \leq t \leq u} \left| \frac{t}{R(t\mathbf{w}_M)/M} - \frac{t}{G(t)} \right| \\ &= \sup_{\delta \leq t \leq u} \left| \frac{t[G(t) - R(t\mathbf{w}_M)/M]}{G(t)R(t\mathbf{w}_M)/M} \right| \leq \frac{\sup_{\delta \leq t \leq u} |G(t) - R(t\mathbf{w}_M)/M|}{G(\delta)R(\delta\mathbf{w}_M)/M} \\ &\rightarrow \frac{0}{G(\delta)^2} = 0 \end{aligned}$$

almost surely, where the numerator converges to 0 by the Glivenko-Cantelli Theorem and the denominator converges to $G(\delta)^2$ by (A4) and the Continuous Mapping Theorem.

As for the second claim, denote $\hat{a}_{0,M}^\lambda = \hat{M}_0(\lambda_M \mathbf{w}_M)/M$ and $a_{0,\infty}^\lambda = [1 - G(\lambda)]/[1 - \lambda]$. Additionally observe

$$\widehat{FDP}_M^\lambda(t\mathbf{w}_M) = \hat{a}_{0,M}^\lambda \widehat{FDP}_M^0(t\mathbf{w}) \quad \text{and} \quad FDP_\infty^\lambda(t) = a_{0,\infty}^\lambda FDP_\infty^0(t),$$

Then using the triangle inequality

$$\begin{aligned} \sup_{\delta \leq t \leq u} \left| \widehat{FDP}_M^\lambda(t\mathbf{w}_M) - FDP_\infty^\lambda(t) \right| &= \sup_{\delta \leq t \leq u} \left| \hat{a}_{0,M}^\lambda \widehat{FDP}_M^0(t\mathbf{w}_M) - a_{0,\infty}^\lambda FDP_\infty^0(t) \right| \\ &\leq |\hat{a}_{0,M}^\lambda - a_{0,\infty}^\lambda| \times \sup_{\delta \leq t \leq u} \left| \widehat{FDP}_M^0(t\mathbf{w}_M) \right| \\ &\quad + a_{0,\infty}^\lambda \times \sup_{\delta \leq t \leq u} \left| \widehat{FDP}_M^0(t\mathbf{w}_M) - \widehat{FDP}_\infty^0(t) \right| \\ &< 2\epsilon + \epsilon, \end{aligned}$$

where the last inequality is satisfied for all large enough M for any $\epsilon > 0$. To verify the last inequality note that $\hat{a}_{0,M}^\lambda \rightarrow a_{0,\infty}^\lambda$ almost surely by (A2), (A4) and the Continuous Mapping Theorem, and hence $|\hat{a}_{0,M}^\lambda - a_{0,\infty}^\lambda| < \epsilon$ for all large enough M . Further, for all large enough M ,

$$\sup_{\delta \leq t \leq u} \widehat{FDP}_M^0(t\mathbf{w}_M) < \sup_{\delta \leq t \leq u} FDP_\infty^0(t) + \epsilon \leq 2$$

by the first claim of the Lemma and (A6). Additionally, $G(\lambda) \geq \lambda$ by (A6) and consequently $a_{0,\infty}^\lambda \leq 1$. Lastly, $\sup_{\delta \leq t \leq u} |\widehat{FDP}_M^0(t\mathbf{w}_M) - FDP_\infty^0(t)| < \epsilon$ for all large enough M by the first claim of the Lemma.

To prove the third claim, we first show that

$$\begin{aligned} & \sup_{\delta \leq t \leq u} |FDP_M(t\mathbf{w}_M) - FDP_\infty(t)| \\ & \leq \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{R(t\mathbf{w}_M)} - \frac{a_0\mu_0 t}{R(t\mathbf{w}_M)/M} \right| + \sup_{\delta \leq t \leq u} \left| \frac{a_0\mu_0 t}{R(t\mathbf{w}_M)/M} - \frac{a_0\mu_0 t}{G(t)} \right| \\ & = \sup_{\delta \leq t \leq u} \frac{M}{R(t\mathbf{w}_M)} \left| \frac{V(t\mathbf{w}_M)}{M} - a_0\mu_0 t \right| \end{aligned} \quad (S7)$$

$$+ a_0\mu_0 \sup_{\delta \leq t \leq u} \left| \widehat{FDP}_M^0(t\mathbf{w}_M) - FDP_\infty^0(t) \right|. \quad (S8)$$

The inequality is a consequence of the triangle inequality and the definitions of $FDP_\infty(t)$ and $FDP_M(t\mathbf{w}_M)$. The expression in (S7) is verified by factoring out $R(t\mathbf{w}_M)/M$ in the first expression on the previous line while the expression in (S8) follows from factoring out $a_0\mu_0$ in the second expression and by the definitions of $\widehat{FDP}_M^0(t\mathbf{w}_M)$ and $FDP_\infty^0(t)$. Now, the quantity in (S8) converges to 0 almost surely because $a_0\mu_0$ is bounded under (A5) and by the first claim of the Lemma. To show that the first expression converges to 0 almost surely, first note for any $t \in (\delta, u]$, because $R(t\mathbf{w}_M)$ is nondecreasing in t , $R(t\mathbf{w}_M)/M > G(\delta/2)$ and hence that

$$\frac{M}{R(t\mathbf{w}_M)} < \frac{1}{G(\delta/2)}$$

for all large enough M . Hence, if

$$\sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M} - a_0\mu_0 t \right| \rightarrow 0 \quad (S9)$$

almost surely, then

$$\sup_{\delta \leq t \leq u} \frac{M}{R(t\mathbf{w}_M)} \left| \frac{V(t\mathbf{w}_M)}{M} - a_0\mu_0 t \right| \leq \frac{\epsilon}{G(\delta/2)}$$

for all large enough M and the proof would be completed since ϵ is arbitrary and δ is fixed. To show (S9), first observe that $E[V(t\mathbf{w}_M)]/M_0 = \bar{w}_{0,M}t$ under the NS conditions. Also note that by the triangle inequality

$$\begin{aligned} \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} - \mu_0 t \right| & \leq \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} - \bar{w}_{0,M}t \right| + \sup_{\delta \leq t \leq u} |\bar{w}_{0,M}t - \mu_0 t| \\ & \leq \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} - \bar{w}_{0,M}t \right| + u |\bar{w}_{0,M} - \mu_0| \rightarrow 0 \end{aligned}$$

almost surely, where the first quantity converges to 0 by the Glivenko-Cantelli Theorem and the second quantity converges to 0 because $\bar{w}_{0,M} \rightarrow \mu_0$ almost surely under (A5) and because $u \leq 1$. Thus, again using the triangle inequality

$$\begin{aligned} \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M} - a_0\mu_0 t \right| & = \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} \left[\frac{M_0}{M} + a_0 - a_0 \right] - a_0\mu_0 t \right| \\ & \leq \left| \frac{M_0}{M} - a_0 \right| \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} \right| + a_0 \sup_{\delta \leq t \leq u} \left| \frac{V(t\mathbf{w}_M)}{M_0} - \mu_0 t \right| \rightarrow 0 \end{aligned}$$

almost surely, where the first quantity converges to 0 because $M_0/M \rightarrow a_0$ almost surely under (A5) and because $V(t\mathbf{w}_M)/M_0 \leq 1$, while the second quantity converges to 0 because $a_0 \leq 1$ and $V(t\mathbf{w}_M)/M_0 \rightarrow \mu_0 t$. Hence we have established (S9). \square

PROOF OF THEOREM 4. Let us first focus on the equalities. Suppose that $t_{\alpha,\infty}^0 < u$. Then $FDP_\infty^0(t_{\alpha,\infty}^0) = \alpha$ by the definition of $t_{\alpha,\infty}^0$ and by (A6). Additionally due to (A6), for any $\epsilon > 0$ there exists a $0 < \delta < \epsilon$ such that

$$FDP_\infty^0(t_{\alpha,\infty}^0 + \delta) < \alpha + \epsilon.$$

Now, Lemma S1 gives $\widehat{FDP}_M^0(t\mathbf{w}_M) < FDP_\infty^0(t_{\alpha,\infty}^0 + \delta)$ for $0 \leq t < t_{\alpha,\infty}^0 + \delta$ and all large enough M . Hence, this and (A6) imply

$$\hat{t}_{\alpha,M}^0 = \sup \left[0 \leq t \leq u : \widehat{FDP}_M^0(t\mathbf{w}_M) \leq \alpha \right] \leq t_{\alpha,\infty}^0 + \delta < t_{\alpha,\infty}^0 + \epsilon$$

for all large enough M . Similar arguments give $\hat{t}_{\alpha,M}^0 > t_{\alpha,\infty}^0 - \epsilon$ for all large enough M . Now if $t_{\alpha,\infty}^0 = u$ then

$$t_{\alpha,\infty}^0 - \epsilon \leq \hat{t}_{\alpha,M}^0 \leq t_{\alpha,\infty}^0 = u$$

for all large enough M . Hence, $|\hat{t}_{\alpha,M}^0 - t_{\alpha,\infty}^0| < \epsilon$ for all large enough M and we conclude $\hat{t}_{\alpha,M}^0 \rightarrow t_{\alpha,\infty}^0$ almost surely. As for the second equality, $FDP_\infty^\lambda(t) = a_{0,\infty}^\lambda FDP_\infty^0(t)$ is also continuous and strictly increasing by (A6) and consequently identical argument apply. Thus $\hat{t}_{\alpha,M}^\lambda \rightarrow t_{\alpha,\infty}^\lambda$ almost surely.

As for the inequality, note that (A6) implies $\lambda \leq G(\lambda)$ which implies

$$a_{0,\infty}^\lambda = \frac{1 - G(\lambda)}{1 - \lambda} \leq 1. \quad (\text{S10})$$

Hence,

$$FDP_\infty^\lambda(t) = a_{0,\infty}^\lambda FDP_\infty^0(t) \leq FDP_\infty^0(t) \quad (\text{S11})$$

for every $t \in (0, u]$. This, (A6) and the definitions of $FDP_\infty^0(\cdot)$, $t_{\alpha,\infty}^0$ and $t_{\alpha,\infty}^\lambda$ imply $t_{\alpha,\infty}^0 \leq t_{\alpha,\infty}^\lambda$. \square

PROOF OF THEOREM 5. By Lemma S1 and (A6), for $0 < s < t \leq u$

$$\begin{aligned} FDP_M(t\mathbf{w}_M) - FDP_M(s\mathbf{w}_M) &> \\ a_0\mu_0 t/G(t) - a_0\mu_0 s/G(s) - 2 \sup_{0 \leq t \leq u} |FDP_M(t\mathbf{w}_M) - a_0\mu_0 t/G(t)| \\ &\rightarrow a_0\mu_0 [t/G(t) - s/G(s)] > 0 \end{aligned}$$

almost surely. Claim (C1) is then a consequence of Theorem 4 and the Continuous Mapping Theorem. To verify Claims (C2) and (C3), first observe that by the triangle inequality

$$\begin{aligned} &|FDP_M(\hat{t}_{\alpha,M}^\lambda \mathbf{w}_M) - FDP_\infty(t_{\alpha,\infty}^\lambda)| \\ &\leq |FDP_M(\hat{t}_{\alpha,M}^\lambda \mathbf{w}_M) - FDP_\infty(\hat{t}_{\alpha,M}^\lambda)| + |FDP_\infty(\hat{t}_{\alpha,M}^\lambda) - FDP_\infty(t_{\alpha,\infty}^\lambda)|. \end{aligned}$$

The first quantity converges to 0 almost surely by Lemma S1 and the second quantity converges to 0 almost surely by Theorem 4 and the Continuous Mapping Theorem. Hence, $FDP_M(t_{\alpha,M}^\lambda \mathbf{w}_M) \rightarrow FDP_\infty(t_{\alpha,\infty}^\lambda)$ almost surely. Thus to prove Claims (C2) and (C3) it suffices to show that $FDP_\infty(t_{\alpha,\infty}^\lambda) \leq \alpha$ if $\mu_0 \leq 1$, with equality when $G(t)$ is a DU distribution with $\mu_0 = 1$ and $FDP_\infty^\lambda(u) \geq \alpha$. To show this, consider the following:

$$\begin{aligned}
 FDP_\infty(t_{\alpha,\infty}^\lambda) &= a_0 \mu_0 \frac{t_{\alpha,\infty}^\lambda}{G(t_{\alpha,\infty}^\lambda)} \\
 &\leq a_0 \frac{t_{\alpha,\infty}^\lambda}{G(t_{\alpha,\infty}^\lambda)} \\
 &\leq \frac{1 - G(\lambda)}{1 - \lambda} \frac{t_{\alpha,\infty}^\lambda}{G(t_{\alpha,\infty}^\lambda)} \\
 &= FDP_\infty^\lambda(t_{\alpha,\infty}^\lambda) \\
 &\leq \alpha.
 \end{aligned}$$

The first equality is due to the definition of $FDP_\infty(\cdot)$. The first inequality is satisfied when $\mu_0 \leq 1$ and is an equality when $\mu_0 = 1$. As for the second inequality, note that $G(\lambda) \leq a_0 \lambda + 1 - a_0$ when $\mu_0 \leq 1$ and $G(\lambda) = a_0 \lambda + 1 - a_0$ under a DU distribution with $\mu_0 = 1$. Consequently

$$a_0 = \frac{1 - [a_0 \lambda + 1 - a_0]}{1 - \lambda} \leq \frac{1 - G(\lambda)}{1 - \lambda}$$

when $\mu_0 \leq 1$ and the inequality is an equality when G is a DU distribution with $\mu_0 = 1$. The last equality is satisfied by the definition of $FDP_\infty^\lambda(\cdot)$. The last inequality is satisfied by the definition of $t_{\alpha,\infty}^\lambda$ and is an equality when G is a DU distribution with $\mu_0 = 1$ and $FDP_\infty(u) \geq \alpha$ because these conditions imply $FDP_\infty(u) = FDP_\infty^\lambda(u) \geq \alpha$. That is, $FDP_\infty(u)$ is continuous and monotone and takes on value α . Hence, $FDP_\infty(t_{\alpha,\infty}^\lambda) \leq \alpha$ if $\mu_0 \leq 1$ with equality if G is a DU distribution with $\mu_0 = 1$ and $FDP_\infty(u) \geq \alpha$. \square

PROOF OF THEOREM 6. Under the conditions of the Theorem

$$\begin{aligned}
 Cov(W_{m,M}, \theta_{m,M}) &= E[W_{m,M} | \theta_{m,M} = 1] E[\theta_{m,M}] - E[W_{m,M}] E[\theta_{m,M}] \\
 &= E[\theta_{m,M}] (E[W_{m,M} | \theta_{m,M} = 1] - 1).
 \end{aligned}$$

Hence, $Cov(W_{m,M}, \theta_{m,M}) \geq 0$ implies $E[W_{m,M} | \theta_{m,M} = 1] \geq 1$ and consequently $E[W_{m,M} | \theta_{m,M} = 0] \leq 1$, with equality if $Cov(W_{m,M}, \theta_{m,M}) = 0$. Hence $E[\bar{W}_{0,M} | \theta_M \neq \mathbf{1}_M] = \mu_0 \leq 1$ with equality if $Cov(W_{m,M}, \theta_{m,M}) = 0$. The result follows because $\bar{W}_{0,M} \rightarrow \mu_0$ almost surely. \square

PROOF OF COROLLARY 1. Observe that $u = 1$ and $\lambda < 1$ is fixed. Hence (A2) is satisfied and (A4) - (A6) are satisfied by the conditions of the Theorem. Therefore Claim (C1) holds by Theorem 5. Now, additionally note that $\mu_0 = 1$ if

$\mathbf{w}_M = \mathbf{1}_M$ and that $FDP_\infty(1) = a_0 \geq \alpha$ under the conditions of the Theorem. Thus Claims (C2) and (C3) hold by Theorem 5. \square

Before proving Theorem 7 we provide Lemma S2. It will be used to verify that optimal weights are weakly dependent so that decision functions satisfy the weak dependence structure defined in (A4) - (A5). Below, denote $t_0(k) = E[t_m(k/p_m, \gamma_m)]$ and denote $G(t_0(k)) = E[\delta_m(t_m(k/p_m, \gamma_m))]$, where the expectations are taken over all random quantities, i.e. over $(Z_m, \theta_m, p_m, \gamma_m)$ for some fixed $k > 0$. Further, define

$$\widetilde{FDP}_\infty(t_0(k)) = \frac{1 - G(t_0(k))}{1 - t_0(k)} \frac{t_0(k)}{G(t_0(k))}$$

and

$$k^* = \inf\{k : \widetilde{FDP}_\infty(t_0(k)) = \alpha\},$$

and denote

$$\tilde{w}_{m,\infty}(k^*, p_m, \gamma_m) = U_m t_m(k^*/p_m, \gamma_m)/t_0(k^*).$$

Lemma S2. Suppose that $\Pr(p_m \leq 1 - \alpha)$. Under Model 1 and (A1), $k_M^* \rightarrow k^*$ almost surely and

$$\tilde{w}_{m,M}(k_M^*, \mathbf{p}, \gamma) \rightarrow \tilde{w}_{m,\infty}(k^*, p_m, \gamma_m)$$

almost surely.

Proof. Note that $0 < \alpha \leq 1 - p_{(M)}$ with probability 1 so that k_M^* is well defined for $M = 1, 2, \dots$ by Theorem 2. Further, observe that $t_m(k^*/p_m, \gamma_m)$, $m = 1, 2, \dots$ are i.i.d. continuous random variables taking values in $[0, 1]$ under Model 1. Hence, by the Strong Law of Large numbers $\bar{t}_M(k^*, \mathbf{p}, \gamma) \rightarrow t_0(k^*)$ almost surely. Likewise, $\bar{G}_M(\mathbf{t}(k^*, \mathbf{p}, \gamma) \rightarrow G(t_0(k^*))$ almost surely and by the Continuous Mapping Theorem

$$\widetilde{FDP}_M(\mathbf{t}(k^*, \mathbf{p}, \gamma)) \rightarrow \widetilde{FDP}_\infty(t_0(k^*))$$

almost surely. Because further $\widetilde{FDP}_M(\mathbf{t}(k, \mathbf{p}, \gamma))$ and $\widetilde{FDP}_\infty(t_0(k))$ are both continuous in k by (A1), we have from the Continuous Mapping Theorem and the definitions of k_M^* and k^* that $k_M^* \rightarrow k^*$ almost surely. Thus,

$$\begin{aligned} \tilde{w}_{m,M}(k_M^*, \mathbf{p}, \gamma) &= U_m \mathbf{w}_{m,M}(k_M^*, \mathbf{p}, \gamma) \\ &= U_m \frac{t_m(k_M^*/p_m, \gamma_m)}{\bar{t}_M(k_M^*, \mathbf{p}, \gamma)} \\ &\rightarrow U_m \frac{t_m(k^*/p_m, \gamma_m)}{t_0(k^*)} = \tilde{w}_{m,\infty}(k^*, p_m, \gamma_m) \end{aligned}$$

almost surely by the Continuous Mapping Theorem. \square

PROOF OF THEOREM 7. First we verify (A2). Observe $\lambda_M = \bar{t}_M(k_M^*, \mathbf{p}, \gamma) \rightarrow t_0(k^*)$ almost surely by the Strong Law of Large Numbers and the Continuous Mapping Theorem, where recall $t_0(k^*) = E[t_m(k^*/p_m, \gamma_m)]$. Thus, by the definition of $\tilde{w}_{m,M}$

$$\lim_{M \rightarrow \infty} \tilde{w}_{m,M} = \lim_{M \rightarrow \infty} \frac{U_m t_{m,M}(k_M^*/p_m, \gamma_m)}{\bar{t}_M(k_M^*, \mathbf{p}, \gamma)} \leq \frac{1}{t_0(k^*)}$$

almost surely, where the last inequality is due to the Continuous Mapping Theorem, Lemma (S2) and because $U_m t_m(k_M^*, \mathbf{p}, \gamma) \leq 1$ almost surely by construction. That is, (A2) is satisfied with $\lambda = u = 1/t_0(k^*)$.

Before verifying (A4) - (A6) we introduce some notation. Denote

$$G^{k^*}(t) = E[\delta_m(t\tilde{w}_{m,\infty}(k^*, p_m, \gamma_m))]$$

where the expectation is taken over all random quantities, i.e. taken over $(Z_m, \theta_m, p_m, \gamma_m, U_m)$. Further we sometimes suppress \mathbf{p} and γ and write $\tilde{w}_{m,\infty}(k^*) = \tilde{w}_{m,\infty}(k^*, p_m, \gamma_m)$, $\tilde{w}_{m,M}(k^*) = \tilde{w}_{m,M}(k^*, \mathbf{p}, \gamma)$ and $\tilde{\mathbf{w}}_M(k^*) = [\tilde{w}_{m,M}(k^*), m \in \mathcal{M}]$. Further, denote $\tilde{\mathbf{w}}_\infty(k^*) = [\tilde{w}_{m,\infty}(k^*), m \in \mathcal{M}]$.

Now consider (A4). Observe that $\delta_m(t\tilde{w}_{m,\infty}(k^*)), m = 1, 2, \dots$ are i.i.d. Bernoulli random variables with success probability $G^{k^*}(t)$ under Model 1 so that

$$\frac{R(t\tilde{\mathbf{w}}_\infty(k^*))}{M} = \frac{\sum_{m \in \mathcal{M}} \delta_m(t\tilde{w}_{m,\infty}(k^*))}{M} \rightarrow G^{k^*}(t)$$

almost surely by the Strong Law of Large Numbers. Further, by the NS assumptions, Lemma S2, and because $G^{k^*}(t)$ is continuous, we have that for any $\epsilon > 0$ there exists an $\epsilon' > 0$ such that

$$\begin{aligned} \frac{R(t\tilde{\mathbf{w}}_M(k_M^*))}{M} &= \frac{\sum_{m \in \mathcal{M}} \delta_m(t\tilde{w}_{m,M}(k_M^*))}{M} < \frac{\sum_{m \in \mathcal{M}} \delta_m(t[\tilde{w}_{m,\infty}(k^*) + \epsilon'])}{M} \\ &< G^{k^*}(t + t\epsilon') < G^{k^*}(t) + \epsilon \end{aligned}$$

for all large enough M . Similar arguments give

$$\frac{\sum_{m \in \mathcal{M}} \delta_m(t\tilde{w}_{m,M}(k^*))}{M} > G^{k^*}(t) - \epsilon$$

for all large enough M . Hence, $R(t\tilde{\mathbf{w}}_M(k^*))/M \rightarrow G^{k^*}(t)$ almost surely. Then because $k_M^* \rightarrow k^*$ almost surely by Lemma S2, $R(t\tilde{\mathbf{w}}_M(k_M^*))/M \rightarrow G^{k^*}(t)$ almost surely by the Continuous Mapping Theorem.

As for (A5), recall the NS conditions give $E[\delta_m(t_m)|\theta_m = 0] = t_m$. Hence, taking the expectation over all random quantities, we have by the law of iterated expectation

$$E[(1 - \theta_m)\delta_m(t\tilde{w}_{m,\infty}(k^*, p_m, \gamma_m))] = a_0\mu_0 t,$$

where $a_0 = E[1 - \theta_m]$ and $\mu_0 = E[\tilde{w}_{m,\infty}(k^*, p_m, \gamma_m)|\theta_m = 0]$. Then, arguments akin to those in the proof of (A4) give

$$\frac{V(t\tilde{\mathbf{w}}_M(k_M^*))}{M} = \frac{M_0}{M} \frac{V(t\tilde{\mathbf{w}}_M(k_M^*))}{M_0} \rightarrow a_0\mu_0 t$$

almost surely.

For (A6), first observe that $G^{k^*}(t) = a_0\mu_0 t + (1 - a_0)G_1(t)$ for $t \leq u$, where

$$G_1(t) = E[\pi_{\gamma_m}(t\tilde{w}_{m,\infty}(k^*, p_m, \gamma_m))]$$

and the expectation is taken over all random quantities. Clearly $t \mapsto G_1(t)$ is concave and twice differentiable because $t \mapsto \pi_{\gamma_m}(t)$ is twice differentiable almost surely by (A1). To see that $t/G(t) \rightarrow 0$ as $t \downarrow 0$ note that $G'_1(t) \rightarrow \infty$ as $t \downarrow 0$ because $\pi'_{\gamma_m}(t) \rightarrow \infty$ as $t \downarrow 0$ almost surely by (A1). Hence,

$$\frac{t}{G^{k^*}(t)} = \frac{t}{a_0\mu_0 t + (1 - a_0)G_1(t)} \rightarrow 0$$

as $t \downarrow 0$ by an application of Hôpital's rule. Clearly, $\lim_{t \uparrow u} t/G^{k^*}(t) \rightarrow u/G^{k^*}(u)$ because $G^{k^*}(t)$ is continuous. To see that $u/G^{k^*}(u) \leq 1$ we establish the following:

$$\begin{aligned} G^{k^*}(u) &= E[\delta_m(u\tilde{w}_{m,\infty}(k^*))] \\ &= a_0 E[\delta_m(u\tilde{w}_{m,\infty}(k^*)) | \theta_m = 0] \\ &\quad + (1 - a_0) E[\delta_m(u\tilde{w}_{m,\infty}(k^*)) | \theta_m = 1] \\ &= a_0 E[u\tilde{w}_{m,\infty}(k^*)] + (1 - a_0) E[\pi_{\gamma_m}(u\tilde{w}_{m,\infty}(k^*))] \\ &\geq a_0 E[u\tilde{w}_{m,\infty}(k^*)] + (1 - a_0) E[u\tilde{w}_{m,\infty}(k^*)] \\ &= E[u\tilde{w}_{m,\infty}(k^*)] \\ &= u E[\tilde{w}_{m,\infty}(k^*)] = u. \end{aligned}$$

The first equality is by the definition of $G^{k^*}(u)$ while the second equality is due to the law of iterated expectation. The third is a consequence of the definition of $\pi_{\gamma_m}(t)$ and the NS assumptions. The inequality is satisfied because $\pi_{\gamma_m}(t) \geq t$ almost surely for every $t \in [0, 1]$ under (A1). The forth equality is obvious. As for the fifth, recall $E[U_m | p_m, \gamma_m] = 1$, $\tilde{w}_{m,\infty}(k^*) = U_m w_{m,\infty}(k^*)$ and that $E[w_{m,\infty}(k^*)] = 1$. Hence, by the law of iterated expectation $E[\tilde{w}_{m,\infty}(k^*)] = E[w_{m,\infty}(k^*)] = 1$.

To verify that $\mu_0 \leq 1$ we make use of Theorem 6 and write $W_m = w_{m,M}(k_M^*, \mathbf{p}, \gamma)$ and $\tilde{W}_m = U_m W_m$ for short. First let us focus on $Cov(W_m, \theta_m)$. From the law of iterated expectation,

$$Cov(W_m, \theta_m) = E[Cov(W_m, \theta_m | p_m)] + Cov(E[W_m | p_m], p_m). \quad (\text{S12})$$

Observe that

$$\begin{aligned} Cov(W_m, \theta_m | p_m) &= E[W_m \theta_m | p_m] - E[W_m | p_m] E[\theta_m | p_m] \\ &= p_m E[W_m | p_m] - p_m E[W_m | p_m] = 0 \end{aligned}$$

which implies that the first expectation in (S12) is 0. To compute the second expectation, first observe $\pi'_{\gamma_m}(t_m)$ is continuous and strictly decreasing and hence

the solution to $\pi'_{\gamma_m}(t_m) = a$, denoted $t_m(a, \gamma_m)$, is continuous and strictly decreasing in a almost surely by (A1). Hence $t_m(k_M^*/p_m, \gamma_m)$ is strictly increasing and continuous in p_m almost surely. Thus,

$$E[W_m | \mathbf{p}, \gamma] = E \left[M \frac{t_m(k_M^*/p_m, \gamma_m)}{t_m(k_M^*/p_m, \gamma_m) + \sum_{j \neq m} t_j(k_M^*/p_j, \gamma_j)} \middle| \mathbf{p}, \gamma \right]$$

is also increasing in p_m almost surely because the function $x/(x+a)$ for a any positive constant is increasing in x for $x > 0$. This implies $E[W_m | p_m]$ is also increasing in p_m almost surely which implies $\text{Cov}(E[W_m | p_m], p_m) \geq 0$. As for $\tilde{W}_m = U_m W_m$,

$$\text{Cov}(\tilde{W}_m, \theta_m) = E[\text{Cov}(U_m W_m, \theta_m | W_m)] + \text{Cov}(E[U_m W_m | W_m], E[\theta_m | W_m])$$

by the law of iterated expectation. But

$$E[\text{Cov}(U_m W_m, \theta_m | W_m)] = E[W_m \text{Cov}(U_m, \theta_m | W_m)] = 0$$

because $\text{Cov}(U_m, \theta_m | W_m)$ is 0 by construction. Additionally,

$$\text{Cov}(E[U_m W_m | W_m], E[\theta_m | W_m]) = \text{Cov}(W_m, E[\theta_m | W_m]) \geq 0$$

because $\text{Cov}(W_m, \theta_m) \geq 0$. Hence, $\text{Cov}(\tilde{W}_m, \theta_m) \geq 0$ and thus, by Theorem 6, $\mu_0 \leq 1$. \square

PROOF OF THEOREM 8. First recall from the proof of Theorem 7 (where here we take $U_m = 1$ almost surely for every m) that $\lambda_M = \bar{t}_M(k_M^*) \rightarrow t_0(k^*)$. Hence, we have

$$FDP_\infty^\lambda(t) = \frac{1 - G^{k^*}(t_0(k^*))}{1 - t_0(k^*)} \frac{t}{G^{k^*}(t)}.$$

Further observe that because $t/G^{k^*}(t)$ is strictly increasing by (A6), then $t_0(k^*) = t_{\alpha, \infty}^\lambda$ by the definition of $t_{\alpha, \infty}^\lambda$. Hence $\bar{t}_M(k_M^*) \rightarrow t_0(k^*) = t_{\alpha, \infty}^\lambda$ almost surely. \square

PROOF OF COROLLARY 2 First observe that $\text{Cov}(w_{m,M}, \theta_{m,M}) = 0$ and hence $\mu_0 = 1$ by Theorem 6. It therefore suffices to show that (A4) - (A6) are satisfied. But $\delta_m(tw_{m,M})$, $m = 1, 2, \dots$ are i.i.d. Bernoulli random variables under Model 1 and the conditions of the Theorem. Hence, $R(t\mathbf{w}_M)/M \rightarrow G(t)$ for $G(t) = E[\delta_m(t\mathbf{w}_{m,M})]$ almost surely by the Strong Law of Large Numbers and (A4) is satisfied. Likewise $(1 - \theta_{m,M})\delta_m(tw_{m,M})$, $m = 1, 2, \dots$ are i.i.d. random variable with mean $a_0 t$ under the NS assumptions and the conditions of the Theorem. Hence,

$$\frac{V(t\mathbf{w}_M)}{M} = \frac{1}{M} \sum_{m \in \mathcal{M}} (1 - \theta_{m,M})\delta_m(tw_{m,M}) \rightarrow a_0 t$$

almost surely by the Strong Law of Large Numbers and (A5) is satisfied. Condition (A6) is verified using arguments identical to those used in verifying (A6)

in the proof of Theorem 7 with $G^{k^*}(t) = G(t)$ and $w_{m,M} = \tilde{w}_{m,\infty}(k^*)$. \square

PROOF OF COROLLARY 3. Observe that $\pi(\mathbf{t}, \mathbf{p}, \gamma)$ is proportional to $\pi(\mathbf{t}, \mathbf{1}, \gamma)$ and hence the maximization of $\pi(\mathbf{t}, \mathbf{p}, \gamma)$ with respect to \mathbf{t} is free of p_m . Thus $\tilde{w}_{m,M}(k, \mathbf{p}, \gamma)$ is independent of p_m and hence independent of θ_m . The result then follows from Theorems 6 and 7. \square

10.4. Data analysis and approximations

For simplicity and because, as demonstrated in Sections 6 and 7, the WAMDF is robust with respect to weight misspecification, we approximate power functions using a multivariate normal model for \mathbf{Y}_m . First, assuming \mathbf{Y}_m has a Multinomial distribution with probability vector \mathbf{p}_m and sample size n_m , we compute Z-score for $T_m = \mathbf{x}^T \mathbf{Y}_m$ by

$$Z_m = \frac{\mathbf{x}^T \mathbf{Y}_m - n_m \bar{x}}{\sqrt{\mathbf{x}^T \hat{\Sigma} \mathbf{x}}},$$

where $\bar{x} = \frac{1}{5} \mathbf{j}^T \mathbf{x}$ and $\hat{\Sigma} = n_m \{diag(\hat{\mathbf{p}}_m) - \hat{\mathbf{p}}_m \hat{\mathbf{p}}_m^T\}$ for $\hat{\mathbf{p}}_m = \mathbf{Y}_m / n_m$. By the Central Limit Theorem, Slutsky's Theorem, and the Continuous Mapping Theorem, Z_m is asymptotically normal (in n_m) with variance 1 and mean/effect size

$$\gamma_m = \frac{n_m [\mathbf{x}^T \mathbf{p}_m - \bar{x}]}{\sqrt{\mathbf{x}^T \Sigma \mathbf{x}}} = \sqrt{n_m} \frac{\sum_i x_i [p_{im} - 1/5]}{\sqrt{\mathbf{x}^T [diag(\mathbf{p}_m) - \mathbf{p}_m \mathbf{p}_m^T] \mathbf{x}}} \equiv \sqrt{n_m} K(\beta),$$

where $K(\beta)$ is some constant that depends on β (and \mathbf{x}) through

$$p_{im} = \frac{e^{\beta x_i}}{\sum_{i=1}^5 e^{\beta x_i}}.$$

Now, for any specification of $K(\beta)$, each power function is approximated using the power expression in Example 1 with $\gamma_m = \sqrt{n_m} K(\beta)$.

Remark: We should note that results in this paper assume each Z_m is continuous. While a continuous, say Normal, approximation for the distribution of Z_m may be reasonable, Z_m does technically have a discrete distribution in this particular application. One route for handling this discreteness is to transform each test statistic so as to have a continuous distribution using an independently generated auxiliary variate, as in Stevens (1950); Habiger and Peña (2011); Peña, Habiger and Wu (2011), while another route is to simply ignore the discreteness and use the Normal approximation. As in Sun and McLain (2012), we have chosen the latter approach here for simplicity and because, as demonstrated in Section 6, the asymptotically optimal WAMDF is robust with respect to weight misspecification. Intuitively the difference between these two approaches should be negligible when the distributions of the Z_m s are well approximated, which appears to be the case here. See Figure S1 below. However, a

more precise assessment of this phenomenon, though warranted, is beyond the scope of this manuscript.

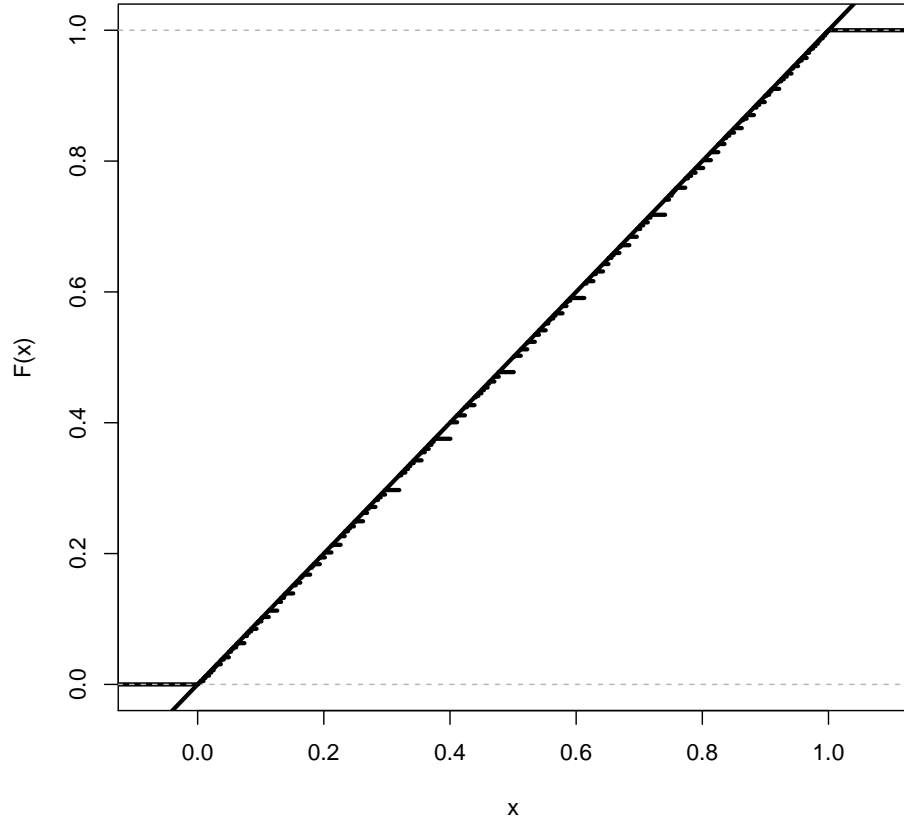


FIG S1. The cumulative distribution function of P_m for $n_m = 6$ under $H_m : \theta_m = 0$ is above. For this data set, $n_m \geq 6$ for each m .